

## A2\_Regresión Múltiple\_fer

Fernanda Pérez

2024-09-17

```
datos <- read.csv("D:/Downloads/AlCorte.csv")
```

```
head(datos)
```

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

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

```
summary(datos)
```

```
##      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
```

```
desviacion_estandar <- apply(datos[, -1], 2, sd)
```

```
varianza <- apply(datos[, -1], 2, var)
```

```
desviacion_estandar
```

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

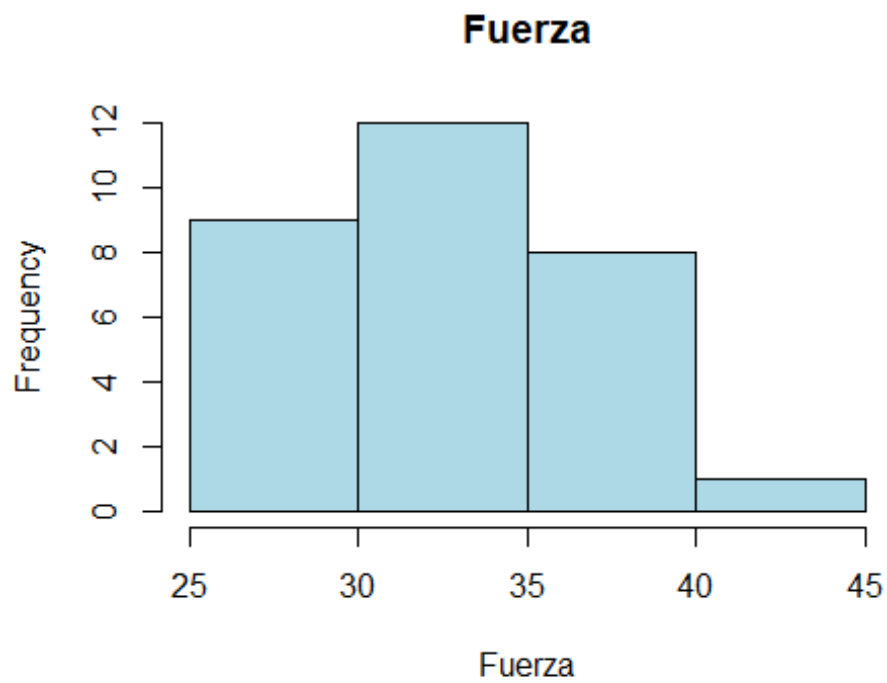
```
varianza
```

```
##      Potencia Temperatura      Tiempo Resistencia
## 186.20690  517.24138  20.68966  80.18133
```

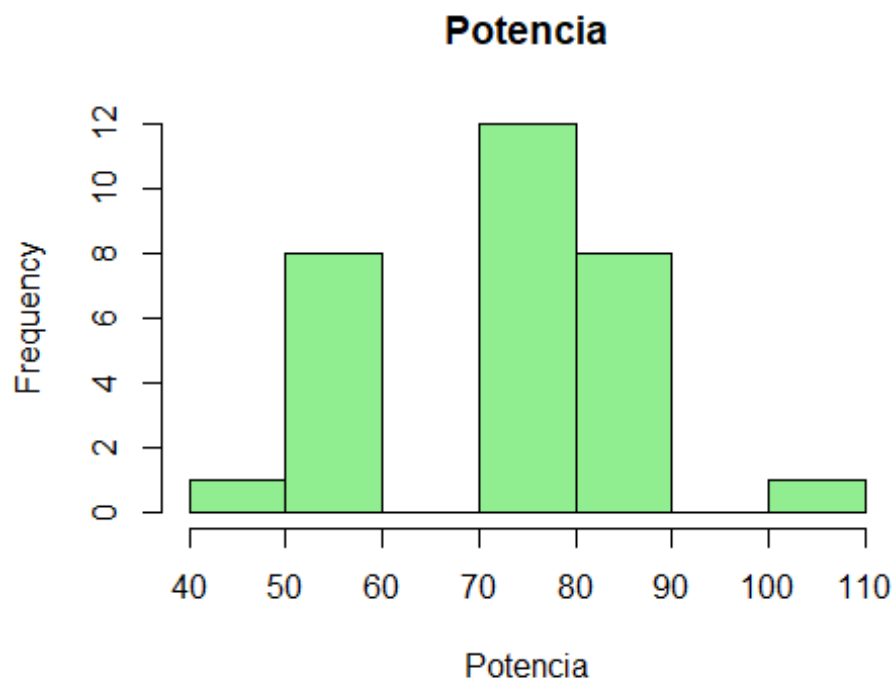
Se observa la desviación estandar de nuestras variables numericas y se calcula la varianza de las variables numericas

### Histogramas de las variables

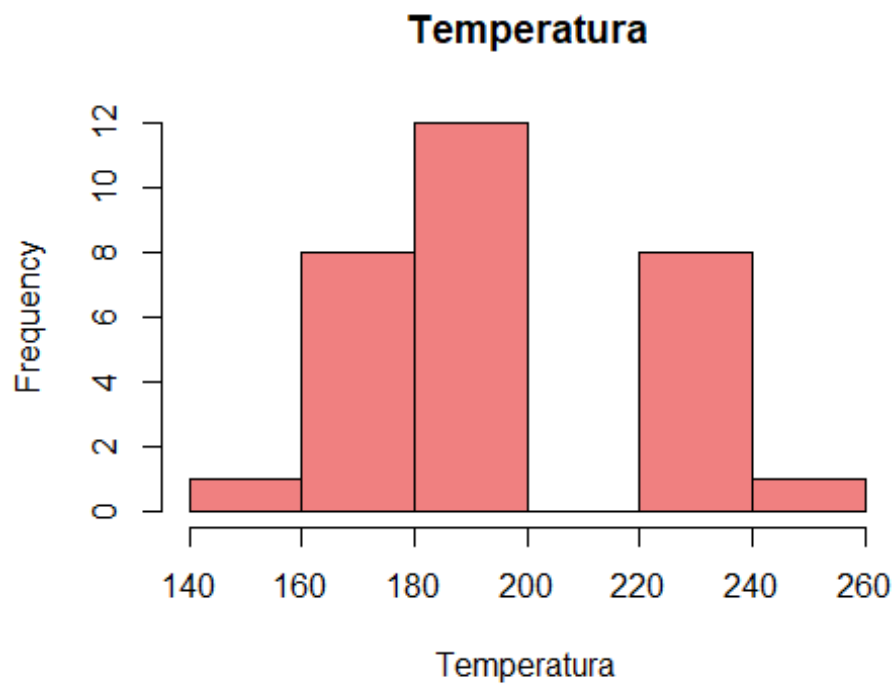
```
hist(datos$Fuerza, main="Fuerza", xlab="Fuerza", col="lightblue")
```



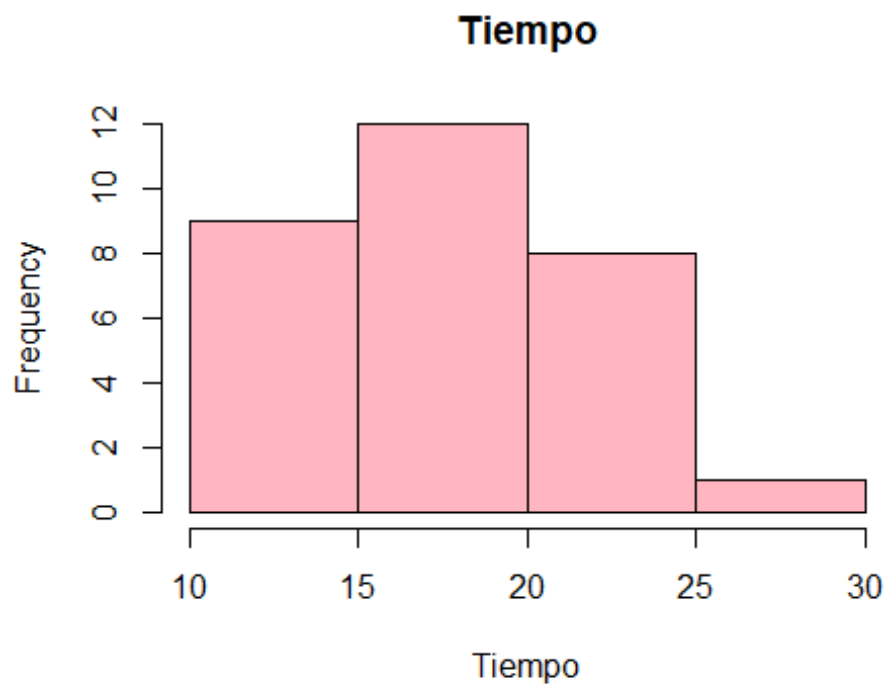
```
hist(datos$Potencia, main="Potencia", xlab="Potencia", col="lightgreen")
```



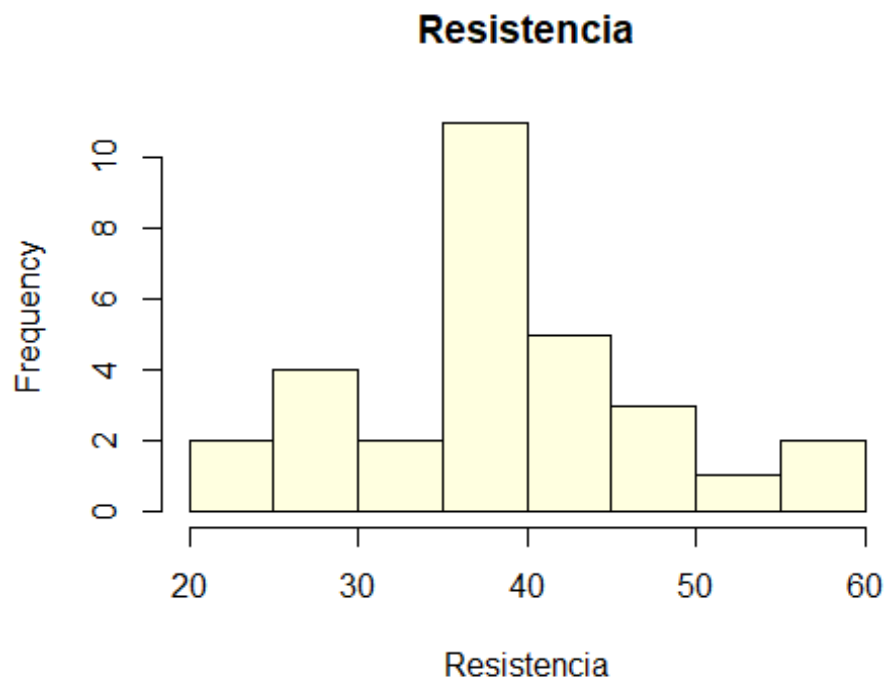
```
hist(datos$Temperatura, main="Temperatura", xlab="Temperatura",  
col="lightcoral")
```



```
hist(datos$Tiempo, main="Tiempo", xlab="Tiempo", col="lightpink")
```



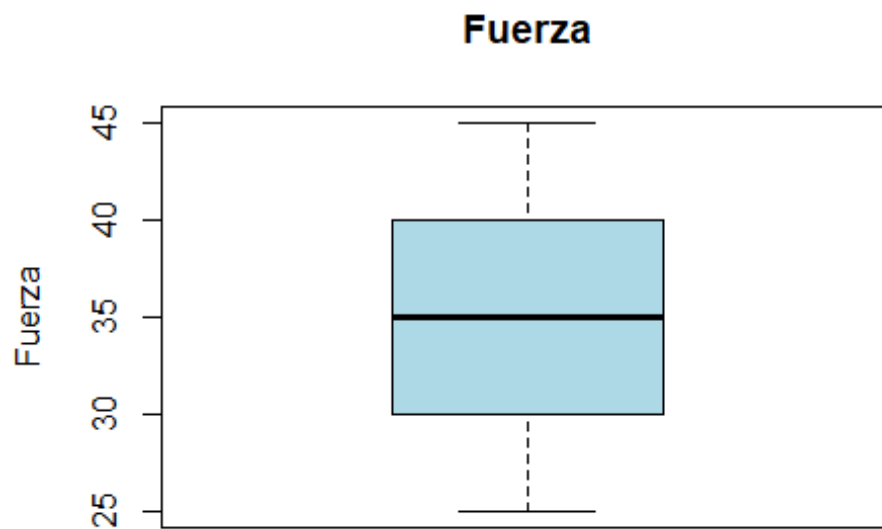
```
hist(datos$Resistencia, main="Resistencia", xlab="Resistencia",  
col="lightyellow")
```



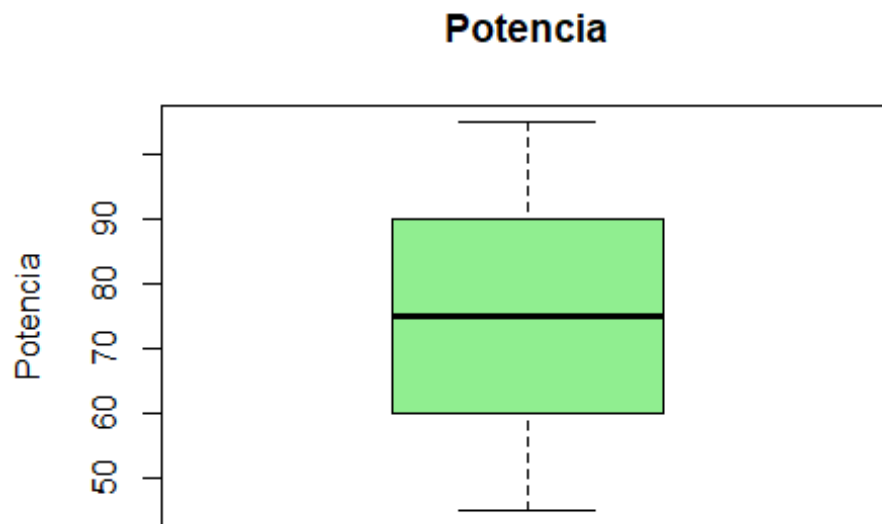
las variables

## Boxplot de

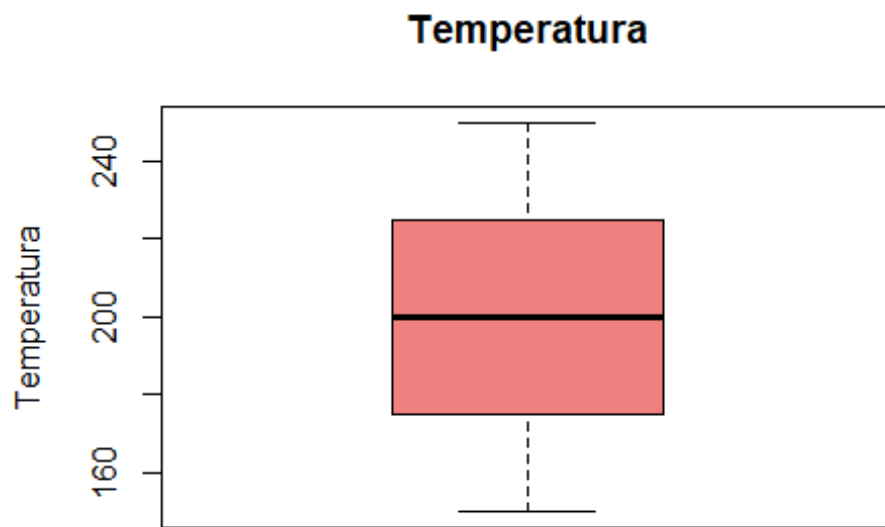
```
boxplot(datos$Fuerza, main="Fuerza", ylab="Fuerza", col="lightblue")
```



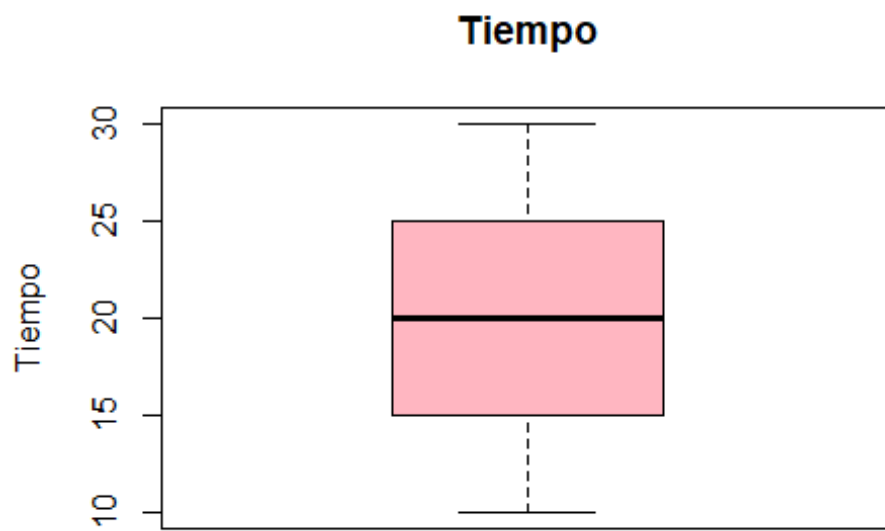
```
boxplot(datos$Potencia, main="Potencia", ylab="Potencia",  
col="lightgreen")
```



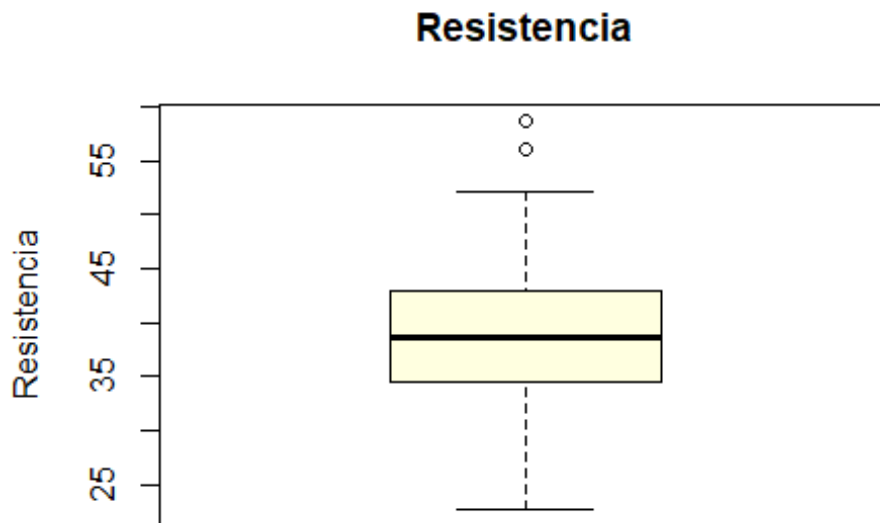
```
boxplot(datos$Temperatura, main="Temperatura", ylab="Temperatura",  
col="lightcoral")
```



```
boxplot(datos$Tiempo, main="Tiempo", ylab="Tiempo", col="lightpink")
```



```
boxplot(datos$Resistencia, main="Resistencia", ylab="Resistencia",  
col="lightyellow")
```



Con estos gráficos vemos la distribución de las variables y la relación con la resistencia al corte. Para darnos una idea general de los datos con los que estamos trabajando.

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

Vamos a tener 3 métodos: mixto, forward y backward; para el modelo completo (cuenta con todas las variables predictoras) y para el modelo nulo (no cuenta con predictores)

```
modeloCompleto <- lm(Resistencia ~ ., data = datos)
```

```
modeloNulo <- lm(Resistencia ~ 1, data = datos)
```

Modelo mixto: se añaden o eliminan variables de acuerdo al criterio de mejor ajuste:

```
pasosMixto <- step(modeloCompleto, direction = "both", trace = 1)
```

```
## Start: AIC=102.96
```

```
## Resistencia ~ Fuerza + Potencia + Temperatura + Tiempo
```

```
##
```

```
##
```

|               | Df | Sum of Sq | RSS    | AIC    |
|---------------|----|-----------|--------|--------|
| - Fuerza      | 1  | 26.88     | 692.00 | 102.15 |
| - Tiempo      | 1  | 40.04     | 705.16 | 102.72 |
| <none>        |    |           | 665.12 | 102.96 |
| - Temperatura | 1  | 252.20    | 917.32 | 110.61 |

```

## - Potencia      1    1341.01 2006.13 134.08
##
## Step:  AIC=102.15
## Resistencia ~ Potencia + Temperatura + Tiempo
##
##              Df Sum of Sq    RSS    AIC
## - Tiempo      1     40.04   732.04 101.84
## <none>                692.00 102.15
## + Fuerza      1     26.88   665.12 102.96
## - Temperatura  1    252.20   944.20 109.47
## - Potencia     1    1341.02 2033.02 132.48
##
## Step:  AIC=101.84
## Resistencia ~ Potencia + Temperatura
##
##              Df Sum of Sq    RSS    AIC
## <none>                732.04 101.84
## + Tiempo      1     40.04   692.00 102.15
## + Fuerza      1     26.88   705.16 102.72
## - Temperatura  1    252.20   984.24 108.72
## - Potencia     1    1341.01 2073.06 131.07

summary(pasosMixto)

##
## Call:
## lm(formula = Resistencia ~ Potencia + Temperatura, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.3233  -2.8067  -0.8483   3.1892   9.4600
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -24.90167    10.07207  -2.472  0.02001 *
## Potencia     0.49833     0.07086   7.033 1.47e-07 ***
## Temperatura  0.12967     0.04251   3.050  0.00508 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.207 on 27 degrees of freedom
## Multiple R-squared:  0.6852, Adjusted R-squared:  0.6619
## F-statistic: 29.38 on 2 and 27 DF,  p-value: 1.674e-07

```

Modelo forward: este metodo se comienza con el modelo nulo y se van añadiendo las variables más significativas

```

pasosForward <- step(modeloNulo, scope = list(lower = modeloNulo, upper =
modeloCompleto), direction = "forward")

```



```

## Start: AIC=132.51
## Resistencia ~ 1
##
##           Df Sum of Sq    RSS    AIC
## + Potencia    1   1341.01   984.24 108.72
## + Temperatura  1    252.20 2073.06 131.07
## <none>                2325.26 132.51
## + Tiempo      1    40.04 2285.22 133.99
## + Fuerza      1    26.88 2298.38 134.16
##
## Step: AIC=108.72
## Resistencia ~ Potencia
##
##           Df Sum of Sq    RSS    AIC
## + Temperatura  1    252.202 732.04 101.84
## <none>                984.24 108.72
## + Tiempo      1    40.042 944.20 109.47
## + Fuerza      1    26.882 957.36 109.89
##
## Step: AIC=101.84
## Resistencia ~ Potencia + Temperatura
##
##           Df Sum of Sq    RSS    AIC
## <none>                732.04 101.84
## + Tiempo  1    40.042 692.00 102.15
## + Fuerza  1    26.882 705.16 102.72

summary(pasosForward)

##
## Call:
## lm(formula = Resistencia ~ Potencia + Temperatura, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.3233  -2.8067  -0.8483   3.1892   9.4600
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -24.90167    10.07207  -2.472  0.02001 *
## Potencia     0.49833     0.07086   7.033 1.47e-07 ***
## Temperatura  0.12967     0.04251   3.050  0.00508 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.207 on 27 degrees of freedom
## Multiple R-squared:  0.6852, Adjusted R-squared:  0.6619
## F-statistic: 29.38 on 2 and 27 DF, p-value: 1.674e-07

```

modelo backward: este metodo se comienza con el metodo completo y se van eliminando las variables significativas

```
pasosBackward <- step(modeloCompleto, 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
##
summary(pasosBackward)

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

De los 3 métodos que tenemos mixto, forward y backward vemos que el modelo final que es más adecuado incluye potencia y temperatura como sus variables predictoras significativas para lograr explicar la resistencia.

Vemos que potencia y temperatura son altamente significativa en todos los modelos ya que en ambos casos  $p < 0.001$ .

## 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

```
anova(pasosMixto)
```

```
## Analysis of Variance Table
##
## Response: Resistencia
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Potencia    1 1341.02  1341.02   49.461 1.465e-07 ***
## Temperatura  1  252.20   252.20    9.302 0.005082 **
## Residuals   27  732.04    27.11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(pasosForward)
```

```
## Analysis of Variance Table
##
## Response: Resistencia
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Potencia    1 1341.02  1341.02   49.461 1.465e-07 ***
## Temperatura  1  252.20   252.20    9.302 0.005082 **
## Residuals   27  732.04    27.11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(pasosBackward)
```

```
## Analysis of Variance Table
##
## Response: Resistencia
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Potencia    1 1341.02  1341.02   49.461 1.465e-07 ***
## Temperatura  1  252.20   252.20    9.302 0.005082 **
```

```
## Residuals    27   732.04    27.11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Vemos que en estos resultados de anova nos aseguramos de demostrar que potencia y temperatura son predictores significativos y robustos para explicar la variabilidad en la Resistencia.

```
summary(pasosMixto)$coefficients
```

```
##              Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) -24.901667  10.07206836 -2.472349 2.001412e-02
## Potencia      0.4983333   0.07085806  7.032839 1.465430e-07
## Temperatura   0.1296667   0.04251483  3.049916 5.082118e-03
```

```
summary(pasosForward)$coefficients
```

```
##              Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) -24.901667  10.07206836 -2.472349 2.001412e-02
## Potencia      0.4983333   0.07085806  7.032839 1.465430e-07
## Temperatura   0.1296667   0.04251483  3.049916 5.082118e-03
```

```
summary(pasosBackward)$coefficients
```

```
##              Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) -24.901667  10.07206836 -2.472349 2.001412e-02
## Potencia      0.4983333   0.07085806  7.032839 1.465430e-07
## Temperatura   0.1296667   0.04251483  3.049916 5.082118e-03
```

Vemos que los coeficientes de los modelos son consistentes y por los interceptos la línea de regresión no pasa por el origen

```
summary(pasosMixto)$r.squared
```

```
## [1] 0.6851783
```

```
summary(pasosMixto)$adj.r.squared
```

```
## [1] 0.6618581
```

```
summary(pasosForward)$r.squared
```

```
## [1] 0.6851783
```

```
summary(pasosForward)$adj.r.squared
```

```
## [1] 0.6618581
```

```
summary(pasosBackward)$r.squared
```

```
## [1] 0.6851783
```

```
summary(pasosBackward)$adj.r.squared
```

```
## [1] 0.6618581
```

Vemos que tenemos  $R^2$  y  $R^2$  ajustados con valores similares y robustos

```
length(coef(pasosMixto)) - 1
```

```
## [1] 2
```

```
length(coef(pasosForward)) - 1
```

```
## [1] 2
```

```
length(coef(pasosBackward)) - 1
```

```
## [1] 2
```

Vemos que independientemente de cual de los 3 modelos elijamos en todos al final se seleccionan 2 variables predictoras: potencia y temperatura.

Vemos que cada modelo acaba con las mismas 2 variables, lo cual nos refleja la importancia de estas 2 variables.

## Analiza la validez del modelo encontrado:

Con los 3 modelos se encontraron resultados muy similares ya que al final todos los modelos terminaron con las mismas 2 variables significativas: potencia y temperatura. El  $R^2$  ajustado también fueron constantemente altos en todos los métodos. Y como al final todos los métodos seleccionaron las mismas variables tienen una economía similar ya sin las variables innecesarias.

Vemos que los métodos tienen un resultado extremadamente similar, es decir son válidos y efectivos, sin embargo el método forward podría ser preferible por el hecho de que por su naturaleza incremental es potencialmente más eficiente en identificar el modelo óptimo haciendo menos cálculos en comparación con el mixto o el backward.

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

```
if (!require(lmtest)) {  
  install.packages("lmtest")  
  library(lmtest)  
}
```

```
## Cargando paquete requerido: lmtest
```

```
## Cargando paquete requerido: zoo
```

```
##
```

```
## Adjuntando el paquete: 'zoo'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## as.Date, as.Date.numeric
```

```

if (!require(car)) {
  install.packages("car")
  library(car)
}

## Cargando paquete requerido: car

## Cargando paquete requerido: carData

modeloForward <- lm(Resistencia ~ Potencia + Temperatura, data = datos)

bptest(modeloForward)

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

dwtest(modeloForward)

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

```

Con estos resultados vemos que cumple adecuadamente con las suposiciones de homocedasticidad e independencia de los residuos. ### No multicolinealidad de Xi

```

vif(modeloForward)

##      Potencia Temperatura
##           1           1

```

Por los valores de VIF vemos que no hay multicolinealidad entre las 2 variables predictoras de nuestro modelo. Un VIF igual a 1 es un valor ideal, y también nos indica que cada predictor es completamente independiente de los otros o sea que está libre de cualquier problema de multicolinealidad.

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

En cada paso se fue realizando un análisis por cada resultado encontrado, después de analizar estos resultados cuidadosamente vemos que el modelo de forward además de ser un muy buen modelo y cumplir con todos los requisitos cumple con las suposiciones básicas de homocedasticidad e independencia de los residuos y está libre de problemas de multicolinealidad, o sea que el modelo forward es estadísticamente

sólido y confiable para la predicción de resistencia al corte basada en las variables de Potencia y Temperatura.