

# Act2 Regresion Multiple

Andrés Villarreal González

2024-09-19

## Regresión Multiple

### Leyendo los datos

```
data <- read.csv("AlCorte.csv")
```

### Medidas principales y gráficos

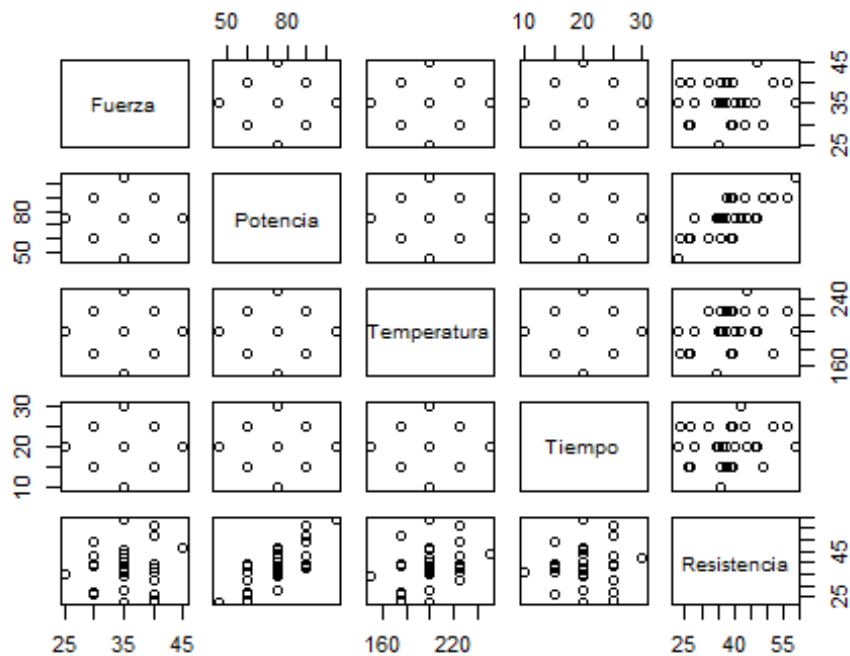
```
summary(data)

##      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

cor(data)

##           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

pairs(data)
```



## Modelo y

analisis del modelo

```
modelo <- lm(Resistencia ~ Fuerza + Potencia + Temperatura + Tiempo, data
= data)
summary(modelo)

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

```

step(modelo, 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
##
## Call:
## lm(formula = Resistencia ~ Potencia + Temperatura, data = data)
##
## Coefficients:
## (Intercept)      Potencia  Temperatura
##    -24.9017         0.4983         0.1297

```

### 1. Significación del modelo global:

El modelo ajustado tiene un valor p global de  $p=1.592 \times 10^{-6}$ , lo que significa que el modelo es estadísticamente significativo. Esto indica que al menos una de las variables predictoras tiene un efecto significativo en la variable de respuesta, Resistencia. Además, el valor de  $R^2$  ajustado es 0.6682, lo que implica que el modelo explica aproximadamente el 67% de la variabilidad total de la resistencia al corte. Este es un porcentaje relativamente alto, lo que sugiere que el modelo tiene un buen ajuste.

### 2. Economía de las variables:

En el modelo inicial, se consideraron cuatro variables: Fuerza, Potencia, Temperatura y Tiempo. Después de realizar la selección de variables mediante el proceso de eliminación hacia atrás (backward elimination), se encontraron que las variables

Fuerza y Tiempo no contribuyen significativamente al modelo y fueron eliminadas. El modelo final incluye solo dos variables predictoras: Potencia y Temperatura. Este modelo es más económico ya que utiliza menos variables, reduciendo el número de parámetros innecesarios, lo que es favorable para evitar el sobreajuste y mejorar la interpretabilidad.

### 3. Significación global (Prueba F):

El valor de la prueba F del modelo inicial fue de 15.6 con un valor p de  $p=1.592 \times 10^{-6}$ . Esto indica que el modelo, en su conjunto, es altamente significativo. Esta prueba evalúa si al menos una de las variables predictoras está relacionada con la variable de respuesta (Resistencia). El valor p extremadamente bajo confirma que el modelo tiene un efecto significativo sobre la resistencia al corte.

### 4. Significación individual:

-Intercepto:  $p=0.00841$ , lo que indica que el intercepto es significativamente diferente de cero.

-Fuerza:  $p=0.32444$ , no significativa, por lo tanto fue eliminada del modelo final.

-Potencia:  $p=1.93 \times 10^{-7}$ , altamente significativa, lo que indica que Potencia es una variable importante para predecir la resistencia al corte.

-Temperatura:  $p=0.00499$ , también significativa, lo que indica que Temperatura tiene un efecto relevante sobre la resistencia.

-Tiempo:  $p=0.23132$ , no significativa, fue eliminada del modelo.

En el modelo final, las dos variables predictoras (Potencia y Temperatura) son estadísticamente significativas, y sus coeficientes tienen valores p menores a 0.05, lo que confirma su relevancia.

### 5. Variación explicada por el modelo:

El R-cuadrado ajustado de 0.6682 indica que el 67% de la variación en la resistencia al corte es explicada por las variables Potencia y Temperatura en el modelo final. Esto significa que el modelo logra capturar una parte significativa de la variabilidad observada en la resistencia al corte, lo que lo hace efectivo para predecir dicha resistencia.

## Nuevo modelo solo con Potencia y Temperatura

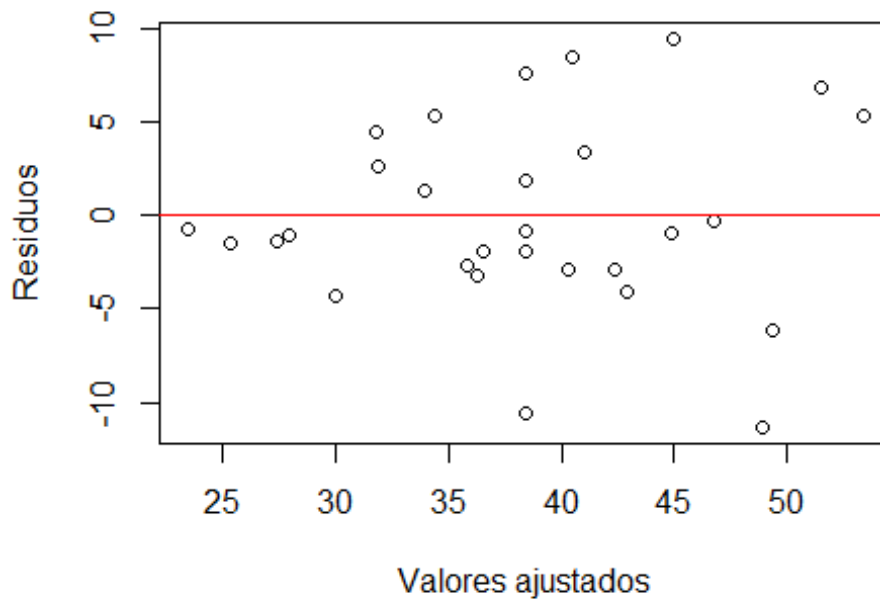
```
modelo2 <- lm(Resistencia ~ Potencia + Temperatura, data = data)
```

## Validez del modelo

### Homocedasticidad

```
plot(modelo2$fitted.values, residuals(modelo2),  
      xlab = "Valores ajustados", ylab = "Residuos",  
      main = "Residuos vs Valores ajustados")  
abline(h = 0, col = "red")
```

## Residuos vs Valores ajustados



```
library(lmtest)

## Loading required package: zoo

##
## Attaching package: 'zoo'

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

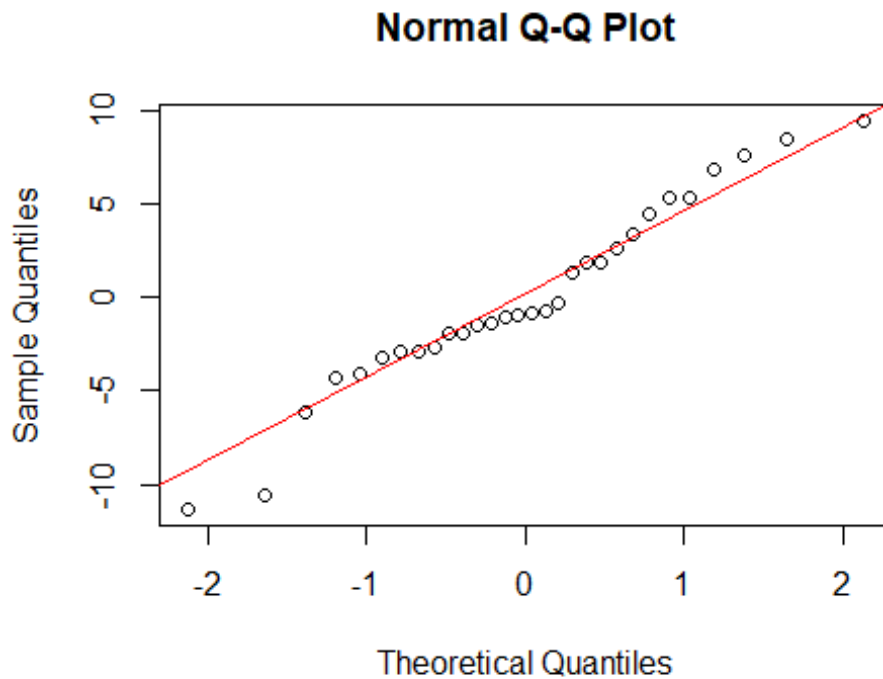
bptest(modelo)

##
## studentized Breusch-Pagan test
##
## data: modelo
## BP = 4.2293, df = 4, p-value = 0.3759
```

Debido al valor p mayor a 0.05 se puede decir que los residuos tienen varianza constante

### Normalidad de los residuos

```
qqnorm(residuals(modelo2))
qqline(residuals(modelo2), col = "red")
```



```
shapiro.test(residuals(modelo2))
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  residuals(modelo2)  
## W = 0.96588, p-value = 0.4333
```

Debido al valor p mayor a 0.05 se puede concluir que los residuos siguen una distribución normal.

#### Independencia

```
library(lmtest)  
dwtest(modelo2)
```

```
##  
##  Durbin-Watson test  
##  
## data:  modelo2  
## DW = 2.3511, p-value = 0.8267  
## alternative hypothesis: true autocorrelation is greater than 0
```

El análisis de la prueba de Durbin-Watson sugiere que los residuos del modelo no tienen problemas de autocorrelación y, por lo tanto, la suposición de independencia de los residuos se cumple en este modelo.

## Conclusion

El modelo encontrado es sólido y bien ajustado. Potencia y Temperatura son las dos variables clave que explican la resistencia al corte, siendo la potencia el factor más relevante. El modelo proporciona una buena comprensión de cómo estas variables influyen en la resistencia, y puede ser utilizado para predecir la resistencia al corte en función de estas dos variables controlables.