

# Estadística y Diseño de Experimentos

## Ejercicios 2

En la fabricación de un motor se deben unir dos tipos de propulsores (tipo 1 y tipo 2). Se sospecha que la resistencia al corte de esta unión está relacionada con la edad (en semanas) del lote de propulsores del tipo 1. En la siguiente tabla se muestra la Resistencia al corte (medida en psi) y la Edad (en semanas) del lote del propulsor tipo 1.

Resistencia	Edad
2158.70	15.50
1678.15	23.75
2316.00	8.00
2061.30	17.00
2207.50	5.50
1708.30	19.00
1784.70	24.00
2575.00	2.50
2357.90	7.50
2256.70	11.00
2165.20	13.00
2399.55	3.75
1779.80	25.00
2336.75	9.75
1765.30	22.00
2053.50	18.00
2414.40	6.00
2200.50	12.50
2654.20	2.00
1753.70	21.50

Cuadro 1: Resistencia obtenida según la edad del propulsor tipo 1.

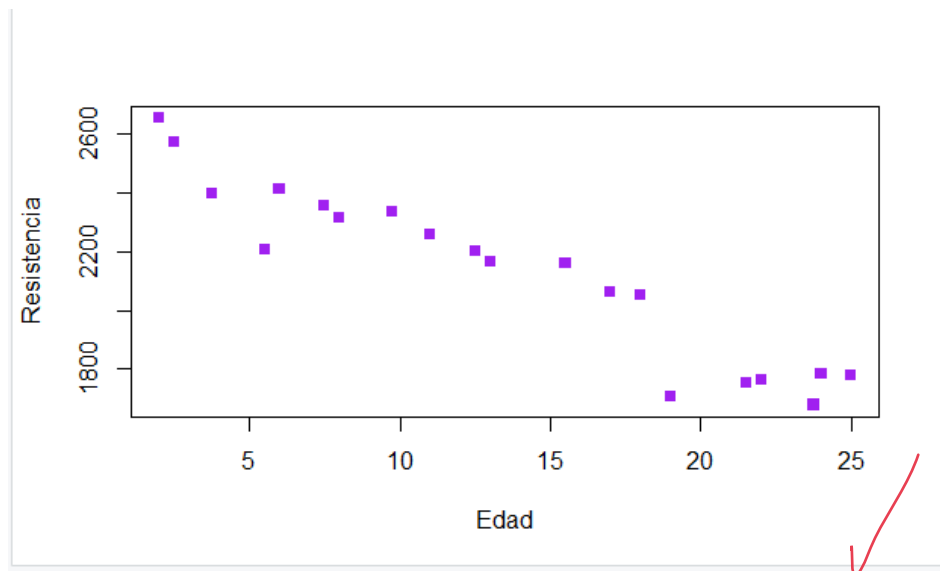
1. Identifique quién es la variable de respuesta Y, quién es la variable regresora X y escriba el modelo de regresión.

- $Y \Rightarrow$  Resistencia al corte
- $X \Rightarrow$  Edad
- Modelo de regresión:

$$Y = \beta_0 + \beta_1$$

$$\text{Resistencia de corte} = \beta_0 + \beta_1 \cdot \text{Edad}$$

## 2. Grafique el diagrama de dispersión de los datos



## 3. Obtenga los estimados para $\beta_0$ , $\beta_1$ y escriba la ecuación de la recta ajustada.

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
+ + + + + Go to file/function Addins
Console Terminal Background Jobs
R - R 4.5.1 - ~/
> #1 Variables X y Y
> Resistencia<-c(2158.70, 1678.15, 2316.00, 2061.30, 2207.50, 1708.30, 1784.70, 2575.00,
+ 2357.90, 2256.70, 2165.20, 2399.55, 1779.80, 2336.75, 1784.70, 2575.00,
+ 2414.40, 2200.50, 2654.20, 1753.70)
> Edad<-c(15.50, 23.75, 8.00, 17.00, 5.50, 19.00, 24.00, 2.50, 7.50, 11.00, 13.00,
+ 3.75, 25.00, 9.75, 22.00, 18.00, 6.00, 12.50, 2.00, 21.50)
>
> #2 Diagrama de dispersión
> plot (Edad,Resistencia,pch=15,col='purple')
>
> #3 Obtenga los estimados para  $\beta_0$ ,  $\beta_1$ 
> lm(Resistencia~Edad)

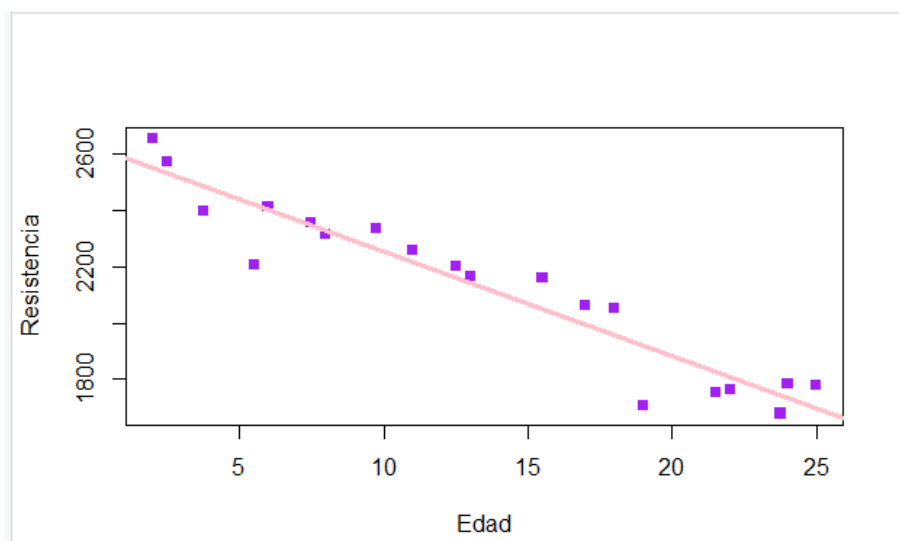
Call:
lm(formula = Resistencia ~ Edad)

Coefficients:
(Intercept)      Edad
    2627.82     -37.15
```

$\beta_0 = 2627.82$  ;  $\beta_1 = -37.15$

Ecuación de la recta ajustada:  $Y = \hat{\beta}_0 + \hat{\beta}_1 * Edad$

4. Grafique la recta de la regresión junto con los datos. ¿Qué tan bueno cree que es el ajuste?



Aunque los datos presentan variabilidad, no se alejan significativamente de la tendencia general, lo que refuerza la confiabilidad del modelo, por lo tanto, es bueno.

5. Efectúe la prueba de significancia de la regresión para un nivel  $\alpha=0.05$ . Escriba el valor del p-valor. ¿Qué conclusiones puede hacer sobre  $\beta_1$ ?

```

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 2627.822    44.184   59.48 < 2e-16 ***
Edad        -37.154     2.889  -12.86 1.64e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 96.11 on 18 degrees of freedom
Multiple R-squared:  0.9018,    Adjusted R-squared:  0.8964
F-statistic: 165.4 on 1 and 18 DF,  p-value: 1.643e-10

```

El p-valor= $1.643e^{-10}$  es mucho más pequeño que  $\alpha=0.05$ , se rechaza  $H_0: \beta_1=0$ , al rechazarla, se puede decir que X (edad) influye en Y (resistencia).

6. Suponga que se tienen tres lotes del propulsor tipo 1, con 5, 10, 15 semanas de edad respectivamente, ¿Cuál es la estimación para la resistencia según el modelo de regresión (para cada lote)?

```
> #6 ¿Cuál es la estimación para la resistencia según el modelo de regresión (para cada lote)?
> predict(RL, list(Edad = c(5, 10, 15)), interval = "confidence")
      fit      lwr      upr
1 2442.054 2374.122 2509.987
2 2256.286 2206.739 2305.834
3 2070.518 2024.289 2116.748
> predict(RL, list(Edad = c(5, 10, 15)), interval = "prediction")
      fit      lwr      upr
1 2442.054 2229.021 2655.087
2 2256.286 2048.385 2464.188
3 2070.518 1863.382 2277.655
>
```

2256.28

Para 5: 2442.054 ; Para 10: 2223.286 ; Para 15: 2070.518

7. Calcule el valor del coeficiente de determinación (Multiple R Squared). Según este coeficiente. ¿Qué tan bueno es el ajuste de la regresión?

```
> #7 Calcule el valor del coeficiente de determinación
> summary(RL)

Call:
lm(formula = Resistencia ~ Edad)

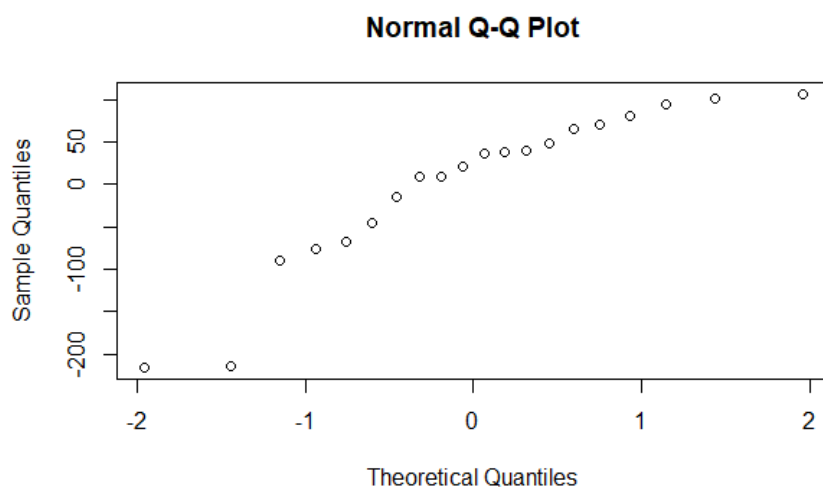
Residuals:
    Min       1Q   Median       3Q      Max
-215.98  -50.68   28.74   66.61  106.76

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2627.822    44.184   59.48  < 2e-16 ***
Edad        -37.154     2.889  -12.86 1.64e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

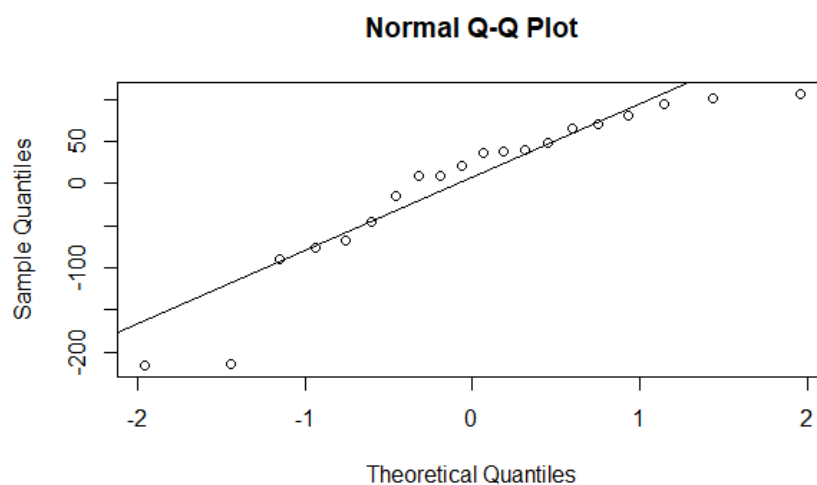
Residual standard error: 96.11 on 18 degrees of freedom
Multiple R-squared:  0.9018,    Adjusted R-squared:  0.8964
F-statistic: 165.4 on 1 and 18 DF,  p-value: 1.643e-10
```

$R^2 = 0.9018$  esto quiere decir que el ajuste del modelo de regresión es **muy bueno**, ya que es cercano a 1 e indica que se ajusta bien a los datos y que la variable edad es un predictor fuerte de la resistencia.

**8. VERIFICACION DE SUPUESTOS DEL MODELO. (Obtenga primero los residuales)**

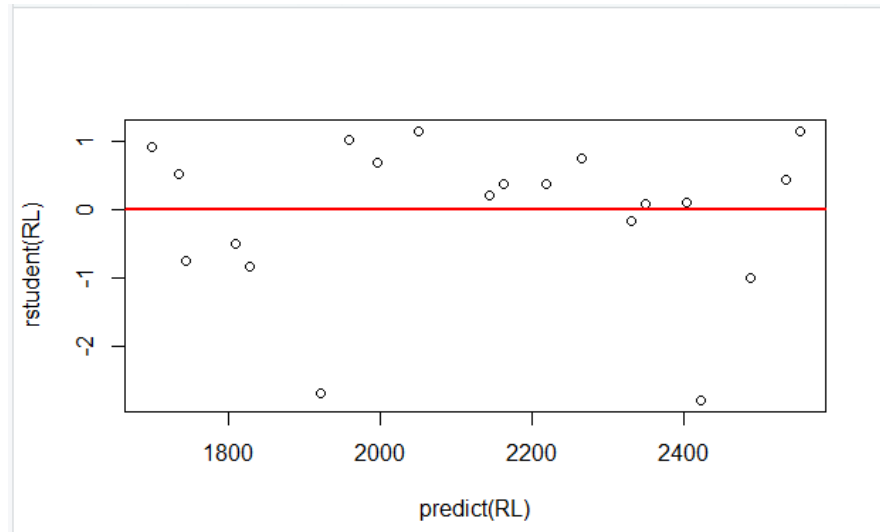


- a. NORMALIDAD. Grafique los residuales contra los cuantiles de una normal (qqnorm, qqline). ¿Se satisface este supuesto?**



El gráfico muestra que la mayoría de los puntos caen sobre la línea diagonal, esto sugiere que la distribución de los errores es aproximadamente “normal”, por lo tanto se satisface.

- b. MEDIA CERO, VARIANZA CONSTANTE E INDEPENDENCIA. Grafique los residuales contra los predichos para verificar los tres supuestos (predict, rstudent). ¿Observa alguna anomalía?**



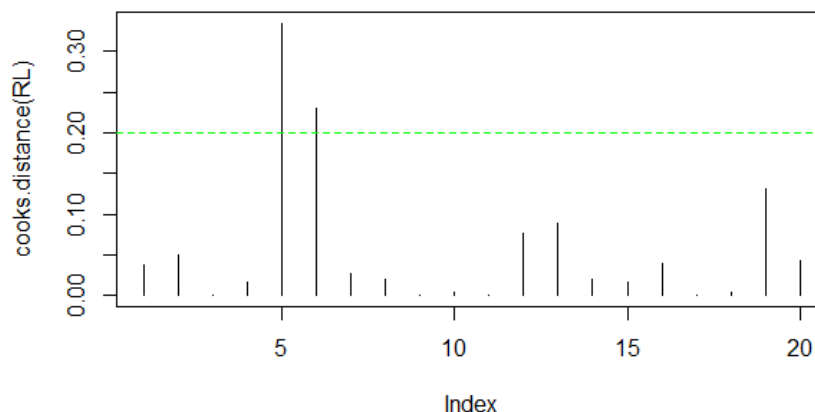
- Media cero: Los residuales están centrados alrededor de cero. Se cumple.
- Varianza constante: No se observa un patrón en forma de abanico o embudo. Los residuales parecen dispersarse por el eje X. Se cumple.
- Independencia: No hay secuencia temporal en los datos, no hay evidencia clara de correlación entre los errores. Se cumple.

## 9. PUNTOS ATÍPICOS E INFLUYENTES.

- a. Utilizando la gráfica anterior, ¿se observan puntos que puedan considerarse como atípicos (outliers)?**

Si hay dos, uno entre 1800-2000 y el otro luego de 2400, se consideran atípicos cuando el rstudent es mayor a  $\pm 2$ .

**b. Utilizando la distancia de Cook, verifique si hay puntos influyentes**



El gráfico permite ver que no hay puntos con distancia de Cook mayor a 1, esto quiere decir que aunque no hay puntos atípicos, no hay evidencia de que sean influyentes.

**10. Escriba una conclusión general para este problema.**

Se cumplen todos los supuestos de normalidad, varianza constante, independencia y media cero. Y aunque hay puntos atípicos, no son influyentes, esto no afecta la validez del modelo, pero entre mayor edad del lote, la resistencia de corte será menor, lo cual es coherente con lo esperado ya que el envejecimiento hace que se reduzca la calidad del material.

## **Código utilizado para Ejercicio 2.**

#1 Variables X y Y.

```
Resistencia<-c(2158.70, 1678.15, 2316.00, 2061.30, 2207.50, 1708.30, 1784.70,  
2575.00,
```

```
2357.90, 2256.70, 2165.20, 2399.55, 1779.80, 2336.75, 1765.30, 2053.50,
```

```
2414.40, 2200.50, 2654.20, 1753.70)
```

```
Edad<-c(15.50, 23.75, 8.00, 17.00, 5.50, 19.00, 24.00, 2.50, 7.50, 11.00, 13.00,
```

```
3.75, 25.00, 9.75, 22.00, 18.00, 6.00, 12.50, 2.00, 21.50)
```

#2 Diagrama de dispersión.

```
plot (Edad,Resistencia,pch=15,col='purple')
```

#3 Obtenga los estimados para  $\beta_0$ ,  $\beta_1$ .

```
lm(Resistencia~Edad)
```

```
RL <- lm(Resistencia~Edad)
```

#4 Recta de regresión

```
abline(RL, col='pink', lwd=3)
```

```
summary(RL)
```

#6 ¿Cuál es la estimación para la resistencia según el modelo de regresión (para cada lote)?

```
predict(RL, list(Edad = c(5, 10, 15)), interval = "confidence")
```

```
predict(RL, list(Edad = c(5, 10, 15)), interval = "prediction")
```

#7 Calcule el valor del coeficiente de determinación

```
summary(RL)
```



#8 Verificación de supuestos del modelo

# a) NORMALIDAD

```
qqnorm(residuals(RL))
```

```
qqline(residuals(RL))
```

# b) MEDIA CERO, VARIANZA CONSTANTE E INDEPENDENCIA.

```
plot(predict(RL), rstudent(RL))
```

```
abline(h=0, col = 'red', lwd = 2)
```

#9 Datos influyentes

```
plot(cooks.distance(RL), type = "h")
```

```
abline(h = 4/20, col = "green", lty = 2)
```