Estadística y Diseño de Experimentos

11 de Octubre, 2025

9

Nombre de lxs alumnxs:

- Diego Núñez Torres

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 esta´ 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.

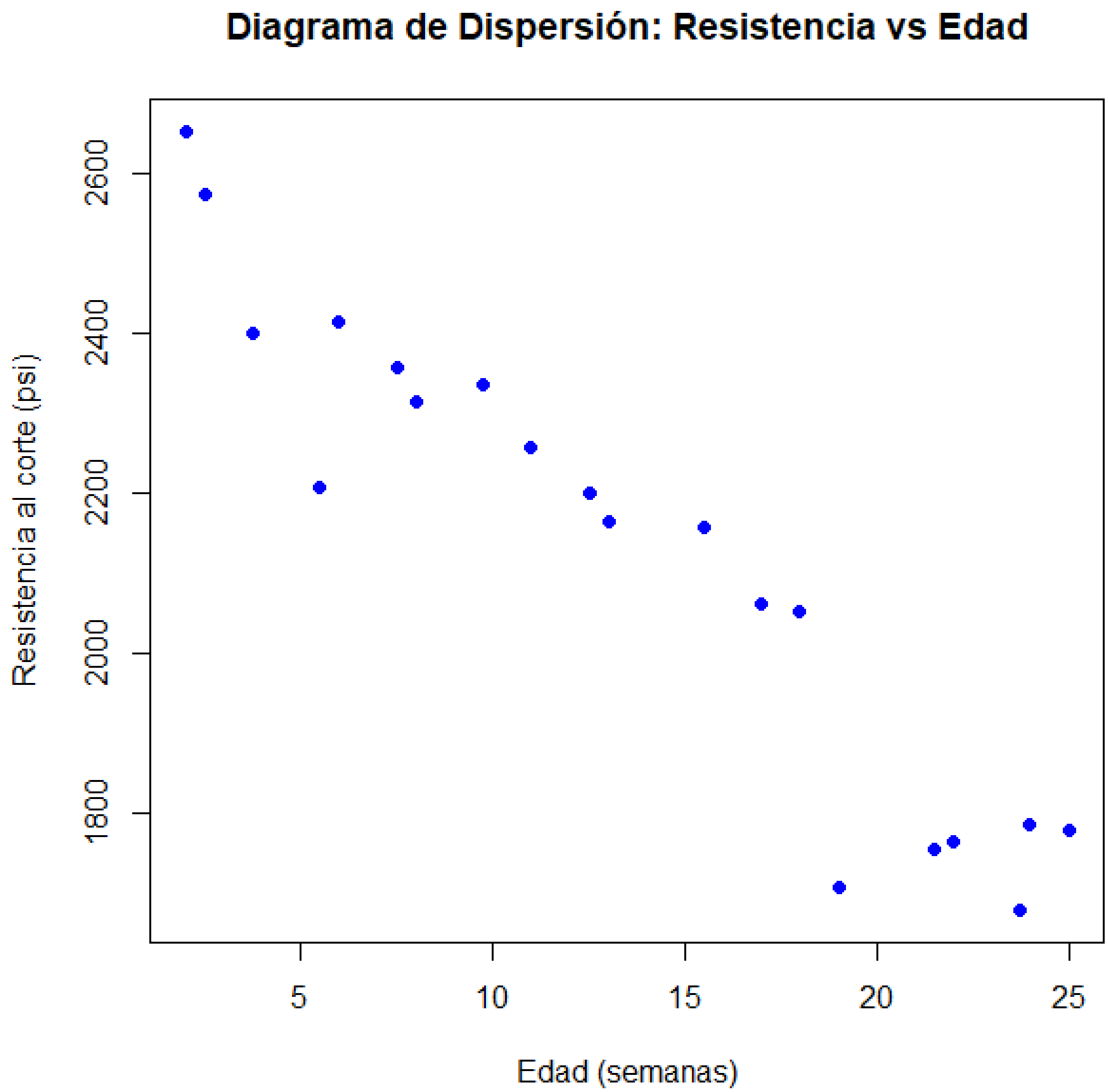
Resistencia	Edad
2158.70	15.50
1678.15	23.75
2316.00	8.00
2061.30	17.00
2207.50	5.50
1708.30	19.00
1784.70	24.00
2575.00	2.50
2357.90	7.50
2256.70	11.00
2165.20	13.00
2399.55	3.75
1779.80	25.00
2336.75	9.75
1765.30	22.00
2053.50	18.00
2414.40	6.00
2200.50	12.50
2654.20	2.00
1753.70	21.50

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

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

   Y = Resistencia (variable de respuesta)
   X = Edad (variable regresora)
   Modelo de Regresión: $Y=\beta 0+\beta 1X+\varepsilon$

2. Grafique el diagrama de dispersión de los datos.



Lo que podemos observar en el diagrama de dispersión es una tendencia lineal negativa, a mayor edad menor es la resistencia al corte.

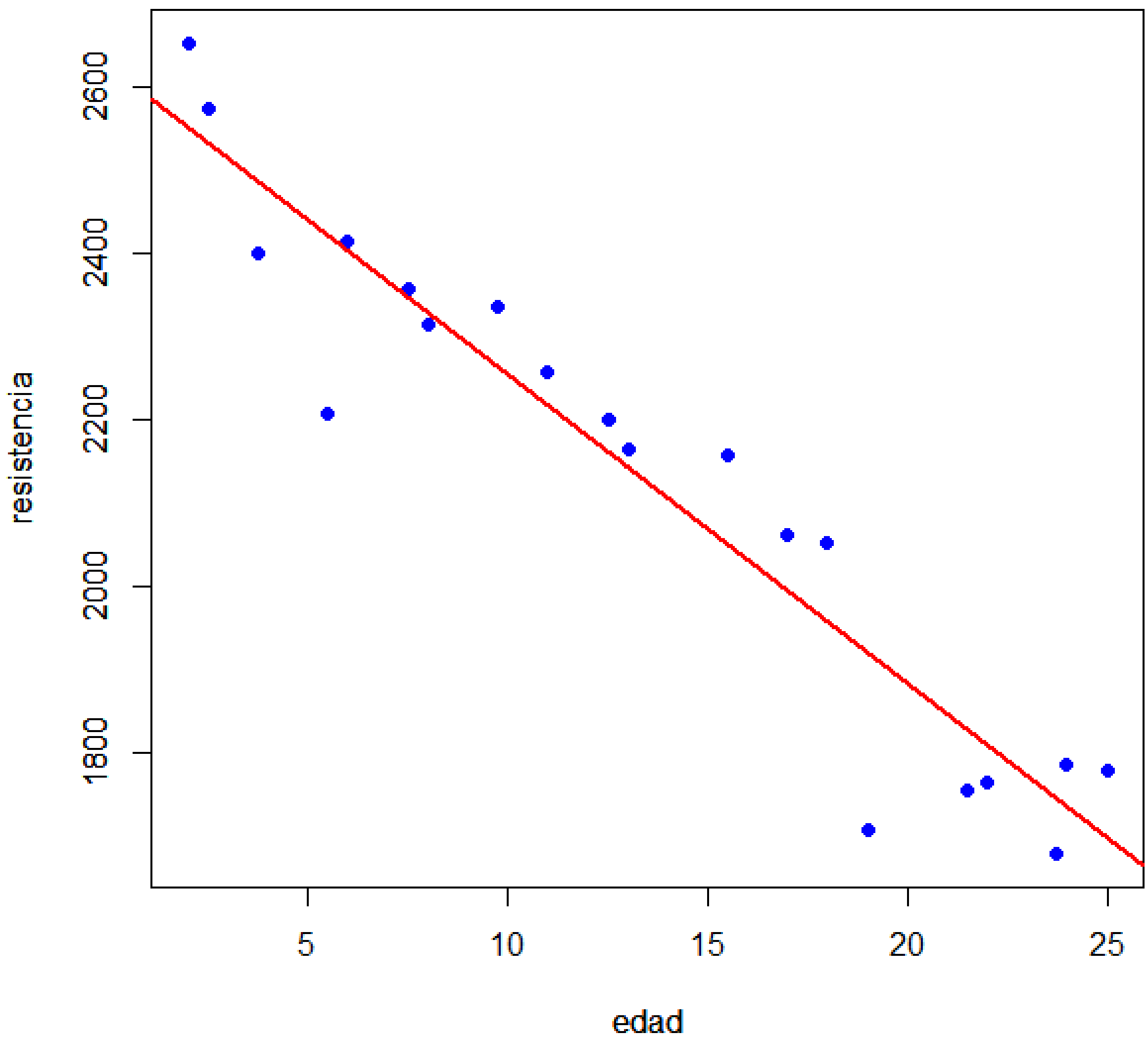
3. Obtenga los estimadores para β_0 y β_1 y escriba la ecuación de la recta ajustada.

Los valores estimados son:

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

Modelo a usar: $\hat{Y} = 2627.82 - 37.15X$

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



La línea roja (modelo usado) tiene una relación alta visualmente, se notan 2 datos ligeramente lejanos de la recta de regresión, por lo que se considera un buen ajuste.

5. Efectúe la prueba de significancia de la regresión para un nivel α =0.05. Escriba el valor del $p - valor$. ¿Qué conclusiones puede hacer sobre β_1?

Para la prueba de significancia, debemos verificar si el coeficiente de β_1 es distinto de 0.

Proponemos hipótesis de prueba:

H0: $\beta_1 = 0$ (La edad no afecta la resistencia)

H1: $\beta_1 \neq 0$ (La edad si afecta la resistencia)

Viendo la parte del código en el inciso 3 (summary), p-valor nos da 1.64e-10, esto lo traducimos como $1.64x10^{-10}$.

Este resultado de p-valor es igual a 0.000000000164, por lo tanto p-valor < α, *porque* 0.000000000164 < 0.05, por lo tanto se rechaza H0, lo cual nos indica que la edad si afecta significativamente la resistencia.

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

Utilizando la función predict en R, la cual nos da los siguientes resultados.

1	2	3
2442.054	2256.286	2070.51

Esto lo traducimos de la siguiente forma:

Edad (en semanas)	Resistencia (psi)
5	2442.054
10	2256.286
15	2070.51

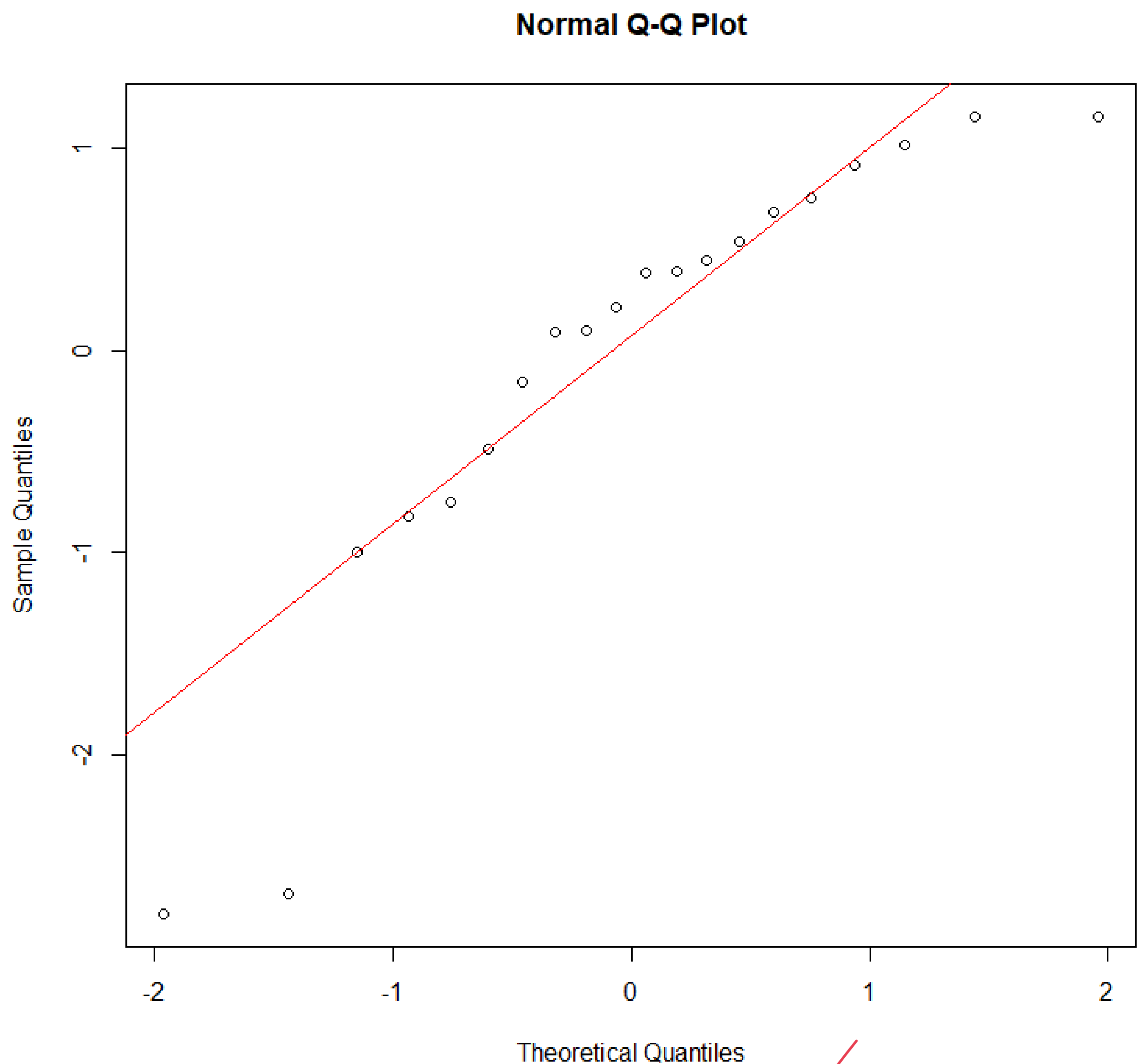
7. Calcule el valor del coeficiente de determinación (Multiple R Squared). Según este coeficiente. ¿qué tan bueno es el ajuste de la regresión?

Usando la función summary en R, editamos para que nos de únicamente el valor deseado, este resultado es: 0.9018414

Este resultado tan cercano a 1, que la relación de la resistencia explicada por la edad es muy alta, por lo que tenemos un excelente ajuste.

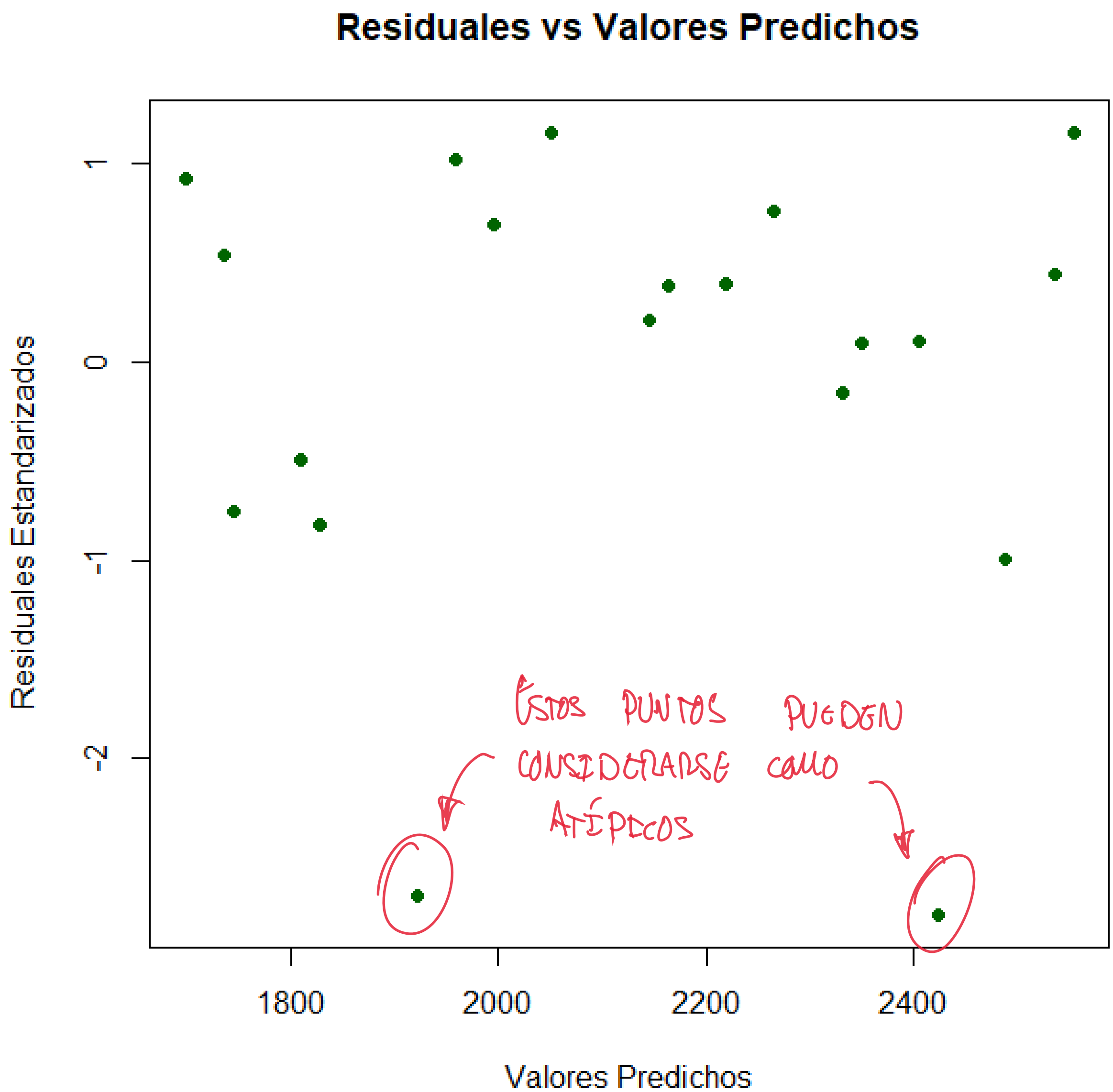
8. VERIFICACION DE SUPUESTOS DEL MODELO. (Obtenga primero los residuales)

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



Se satisface el supuesto, los puntos siguen la línea de forma considerable, se tiene una normalidad razonable.

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



Conforme el gráfico anterior podemos decir que se cumple la media cero, porque los resultados están tanto arriba como debajo de 0 en el eje Y. La varianza constante tiene una dispersión relativamente constante, salvo los 2 puntos de abajo que se separan ligeramente de la mayoría de datos. En cuanto a su independencia, no observamos que los datos sigan una tendencia, por lo que se cumple varianza independiente.

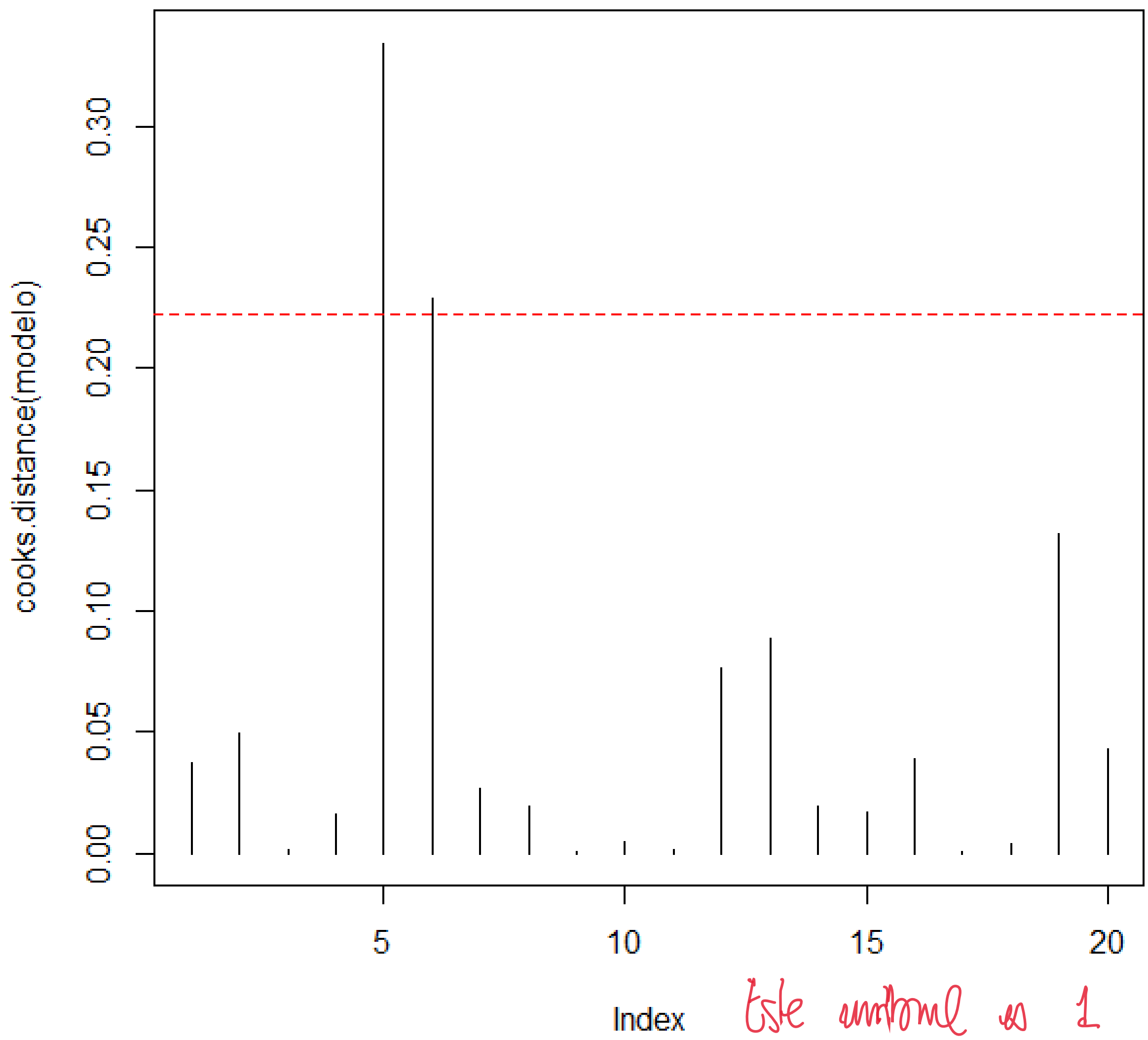
No se observan anomalías, los puntos de abajo que se separan ligeramente de los datos se les podría considerar influyentes, pero no tiene una dispersión grave para ser considerados anomalías.

9. PUNTOS ATÍPICOS E INFLUYENTES.

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

No hay puntos que se puedan considerar atípicos.

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



Este umbral es 1

Existen 2 puntos influyentes, que sobrepasan la línea del umbral común, esto quiere decir que quitar los puntos 5 y 6 afectarían el modelo.

10. Escriba una conclusión general para este problema.

Como lo vimos durante el desarrollo del trabajo, existe una relación lineal negativa significativa entre la edad del propulsor y su resistencia al corte, el modelo obtenido $\hat{Y} = 2627.82 - 37.15X$ explica un 90% de la variabilidad total, lo que significa que nuestro modelo esta bien ajustado, y revisando los supuestos encontramos que no tenemos datos atípicos pero si contamos con datos altamente influyentes en el modelo.

Por lo tanto podemos decir que si la edad del propulsor amento, su resistencia al corte disminuye de forma significativa.

Código usado en R:

```
# ==============================
> # Datos
> 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. Gráfico de dispersión
> plot(edad, resistencia,
+     main = "Diagrama de Dispersión: Resistencia vs Edad",
+     xlab = "Edad (semanas)",
+     ylab = "Resistencia al corte (psi)",
+     pch = 19, col = "blue")
>
> # 3. Ajuste del modelo de regresión lineal
> modelo <- lm(resistencia ~ edad)
> summary(modelo)
```

```
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

>
> # 4. Gráfico con la recta ajustada
> plot(edad, resistencia, pch=19, col="blue")
>abline(modelo, col="red", lwd=2)

> # 6. Predicciones para edades 5, 10 y 15 semanas
> nuevas_edades <- data.frame(edad = c(5, 10, 15))
> predict(modelo, nuevas_edades)
```

```
       1        2        3
2442.054 2256.286 2070.518
>
> # 7. Coeficiente de determinación
> summary(modelo)$r.squared
[1] 0.9018414
>
> # 8. Verificación de supuestos
> residuales <- rstudent(modelo)
> predichos <- predict(modelo)
>
> # a) Normalidad
> qqnorm(residuales)
> qqline(residuales, col = "red")
>
> # b) Media cero, varianza constante e independencia
> plot(predichos, residuales,
+      main = "Residuales vs Valores Predichos",
+      xlab = "Valores Predichos",
+      ylab = "Residuales Estandarizados",
+      pch = 19, col = "darkgreen")
> abline(h = 0, col = "red")
>
> # 9. Puntos atípicos e influyentes
> # a) Outliers visuales
> plot(predichos, residuales, main="Detección de Outliers", xlab="Predichos",
ylab="Residuales")
```

```
> abline(h = 0, col = "red")
>
> # b) Distancia de Cook
plot(cooks.distance(modelo), type="h")
abline(h=4/(20-2), col="red", lty=2)
# ==============================
```