9

Estadística y Diseño de Experimentos
Ejercicios 2

Nombre de lxs alumnxs:

- Leonardo camacho sanchez
-

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* y quién es la variable regresora *X* y escriba el modelo de regresión.

Variable de respuesta Y : es la variable dependiente , resistencia al corte porque es el valor que queremos predecir

Variable regresora: es la variable independiente , edad del propulsor

Modelo:

Regresion lineal simple

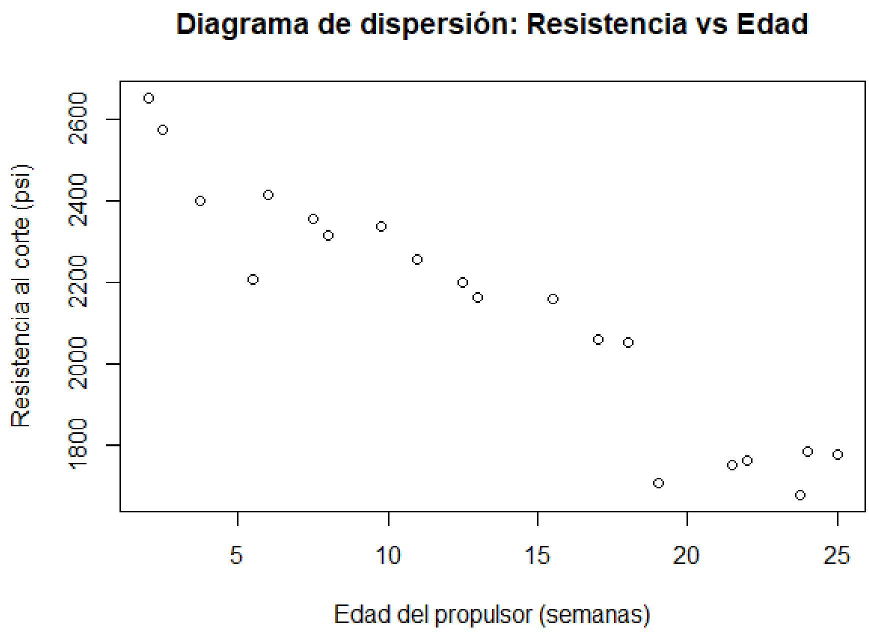
$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

- *Yi* la resistencia al corte del propulsor i
- *Xi* es la edad (en semanas) del propulsor i
- *β0* indica el valor de la resistencia cuando la edad es 0.

¡no siempre es cierta esa interpretación!

- *β1* es la pendiente, que muestra cuánto cambia en promedio la resistencia por cada semana adicionañ
- *εi* es el error aleatorio

2. Grafique el diagrama de dispersión de los datos.



Código en R:

```
plot(datos$Edad, datos$Resistencia,
+     main = "Diagrama de dispersión: Resistencia vs Edad",
+     xlab = "Edad del propulsor (semanas)",
+     ylab = "Resistencia al corte (psi)")
```

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

β_0 = 2610.1

β_1= -36.5

Son ligeramente distintos

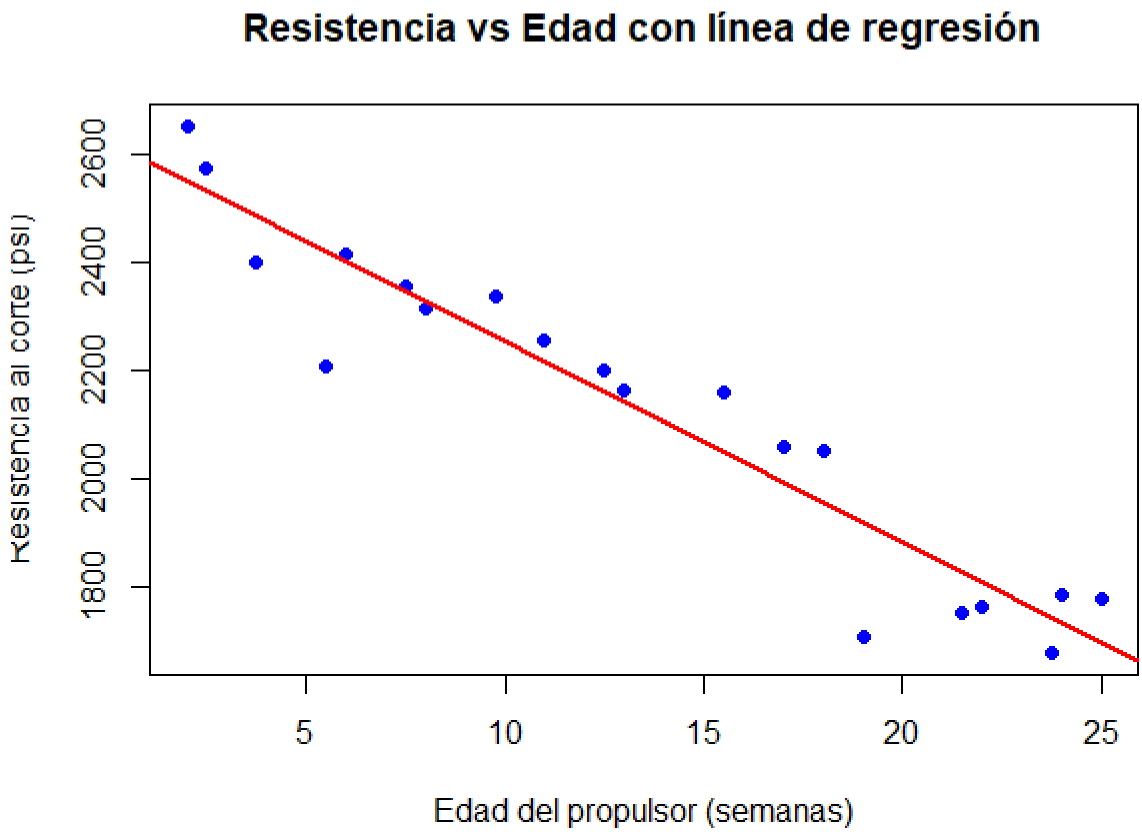
$\hat{\beta_0} \cong 2627.82$

$\hat{\beta_1} \cong -37.1539$

código en R:

```
modelo <- lm(Resistencia ~ Edad, data = datos)
summary(modelo)
```

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



Codigo en R:

```
 plot(datos$Edad, datos$Resistencia,
+    main = "Resistencia vs Edad con línea de regresión",
+    xlab = "Edad del propulsor (semanas)",
+    ylab = "Resistencia al corte (psi)",
+    pch = 19, col = "blue")
>
> abline(modelo, col = "red", lwd = 2)
```

Los punto se distribuyen cerca de la línea lo que indica que el ajuste fue bueno

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?

Nivel de significancia: α = 0.05

Hipótesis:

- H0: β1 = 0 → la edad no afecta la resistencia
- H1: β1 ≠ 0 → la edad sí afecta la resistencia
- 1.643344e-10

p-valor=1.643344×10−10≪0.05

Como el p-valor es mucho menor que 0.05**,** rechazamos H0**.**

La pendiente β1 es diferente de 0, por lo que la edad del propulsor tiene un efecto significativo sobre la resistencia al corte

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)?

Edad Resistencia Estimada

5	2432.8	ligeramente distintos	2442.05
10	2255.5		2256.28
15	2078.2		2070.51

Código en R:

```
nuevas_edades <- data.frame(Edad = c(5, 10, 15))
predicciones <- predict(modelo, newdata = nuevas_edades)
> predicciones
```

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

R-squared: 0.9018

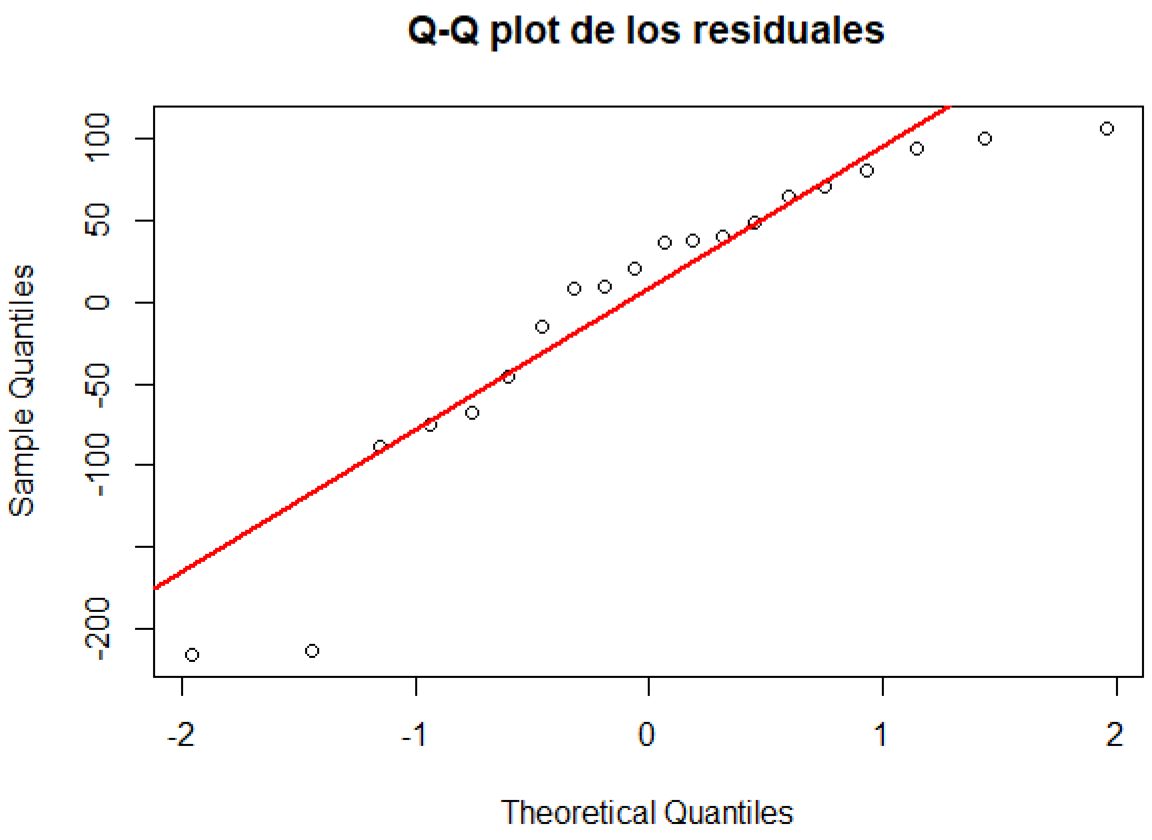
CODIGO EN R

```
summary(modelo)
```

Es muy bueno ya que es un valor cercano a 1

8. VERIFICACIÓN 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?



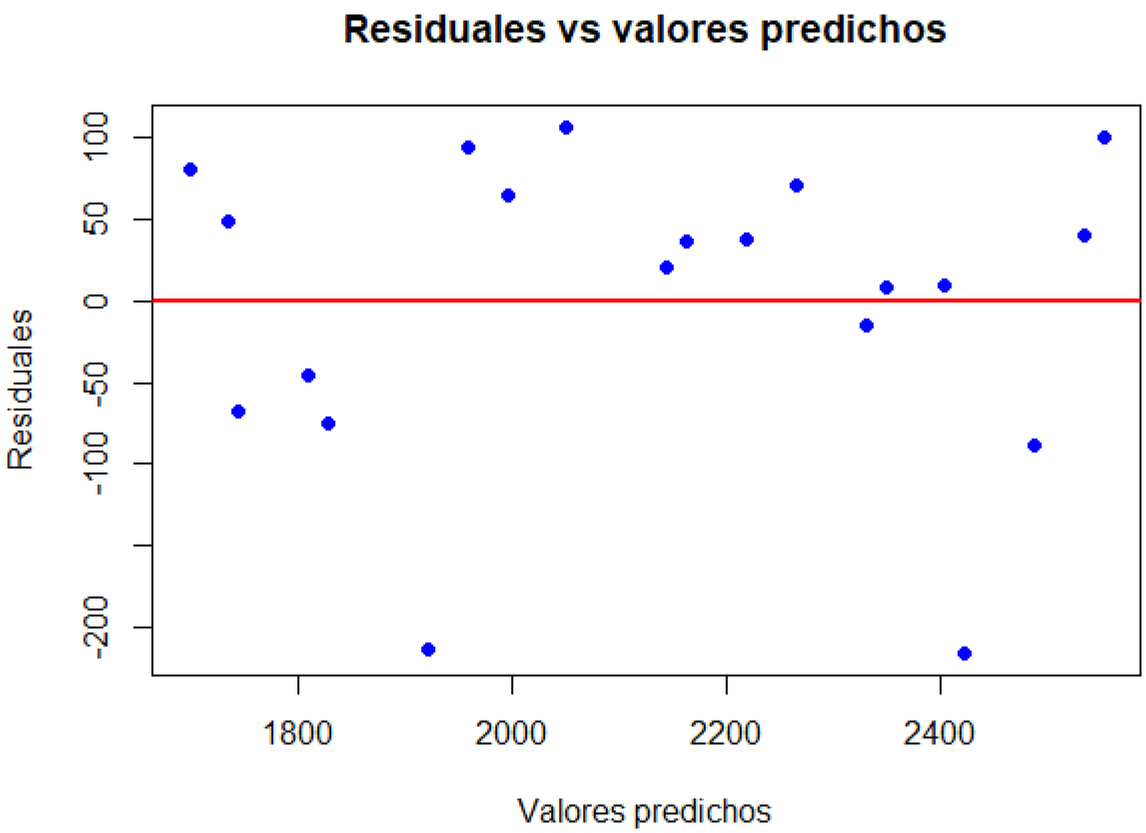
Codigo en R:

```
residuales <- resid(modelo)
qqnorm(residuales, main = "Q-Q plot de los residuales")
```

➢ qqline(residuales, col = "red", lwd = 2

se satisface porque siguen la línea roja

b) MEDIA CERO, VARIANZA CONSTANTE E INDEPENDENCIA. Grafique los re-
siduales contra los predichos para verificar los tres supuestos (predict, rstudent).
¿Observa alguna anomal´ıa?



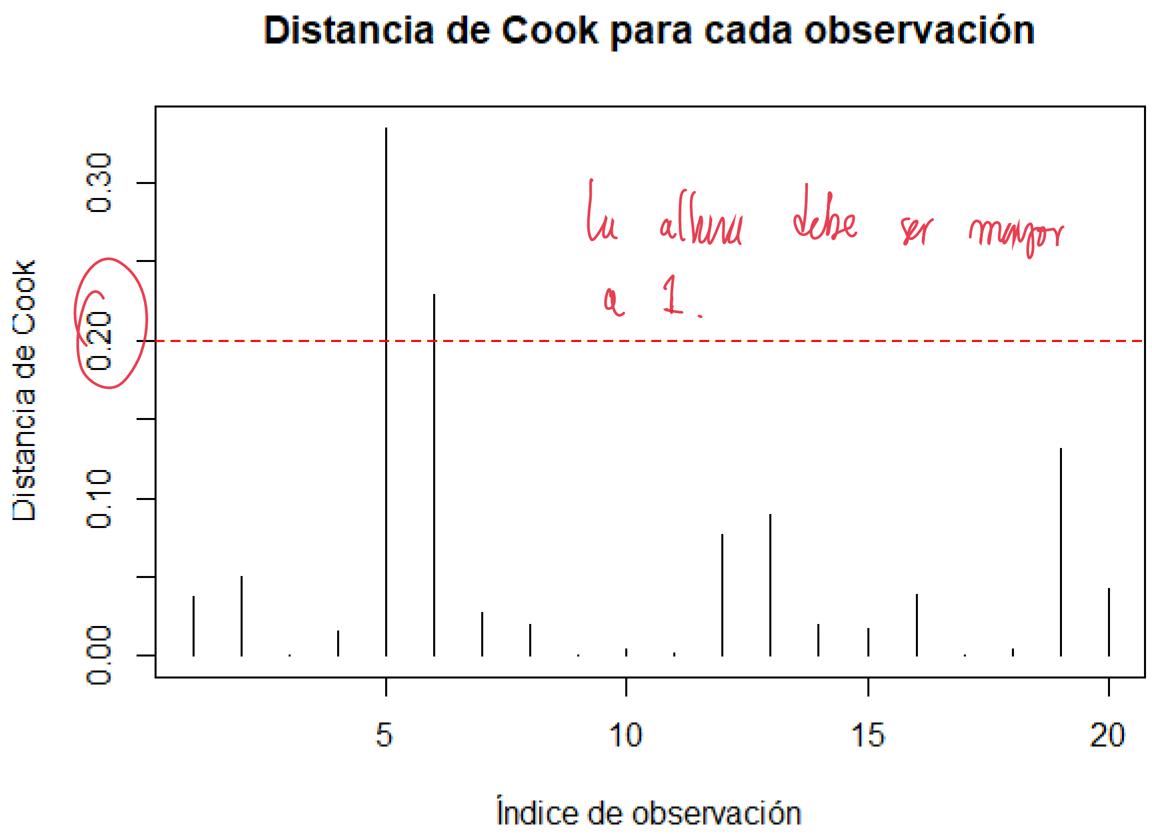
1) Los residuales se distribuyen tanto como por encima como por debajo de la línea
2) Hay algunos valores extremos ¿ Cuáles ?
3) No parece que los punto estan formando algún patron están de forma aleatoria
   Por lo tanto todo normal simplemente esos puntos que están en los valores extremos

9. PUNTOS ATÍPICOS E INFLUYENTES.

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

Si hay algunos puntos que se encuentra en los extremos

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



Se ve que el 5 y 6 estan por encima lo que muestra que son puntos influyentes

Código en R

```
plot(cooksD, type = "h",
+     main = "Distancia de Cook para cada observación",
+     ylab = "Distancia de Cook",
+     xlab = "Índice de observación")
abline(h = 4/length(cooksD), col = "red", lty = 2)
```

10. Escriba una conclusión general para este problema.

El modelo de regresión lineal nos dio una relación línea negativa como vemos en las imágenes en promedio por cada semana adicional la resistencia al corte disminuye aproximadamente 37.15

El modelo tiene un R2= 0.9018 lo que significa que es un buen modelo el 90.18% de la variabilidad

En conclusión, el modelo es correcto para la relación entre la edad y la resistencia al corte