9.5

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

Para esta tabla **la variable de respuesta <u>Y es la resistencia</u>** y la **variable regresora <u>X es la edad</u>**

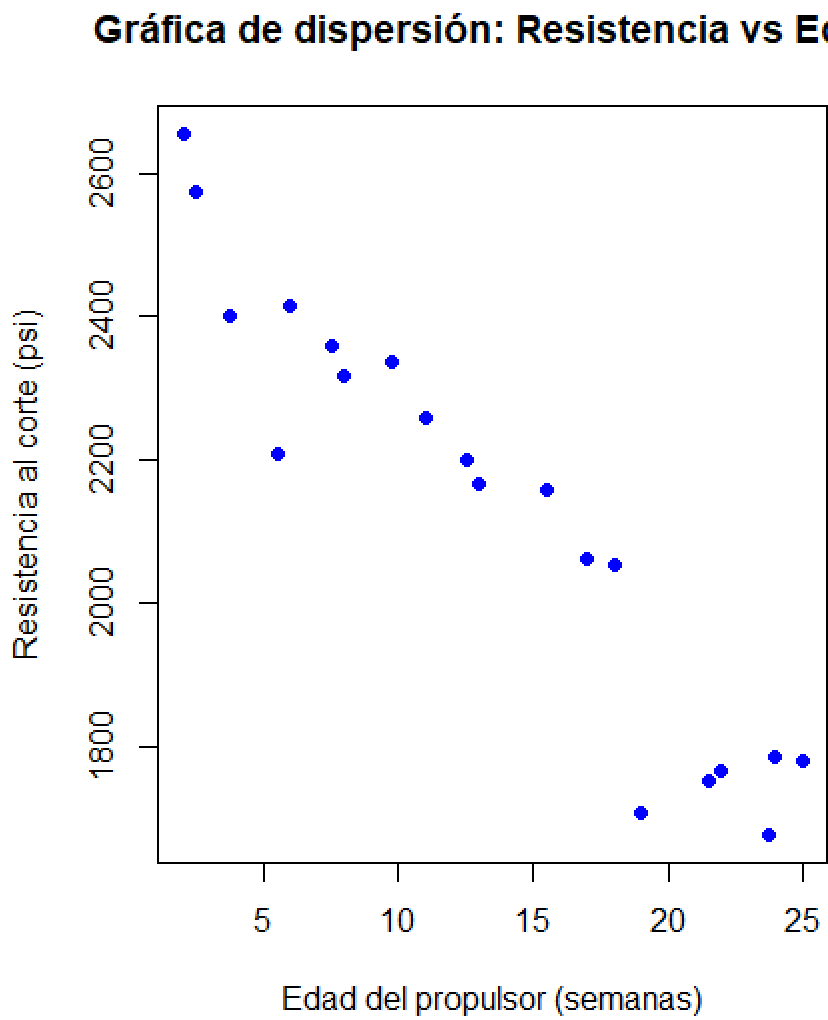
El modelo general de una **regresión lineal simple** es:

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

Aplicando a nuestro problema:

Resistencia=β_0+β_1×Edad$_i$

2. Grafique el diagrama de dispersión de los datos.



3. Obtenga los estimadores para β0 y β1 y escriba la ecuación de la recta ajustada.

Usando R con lm(Resistencia ~ Edad), obtenemos que :

(Intercept) Edad

2627.82236 -37.15359

Lo que significa:

$$\hat{\beta}_0 = 2627.82236 \text{ (intercepto)}$$

$$\hat{\beta}_1 = -37.15359 \text{(pendiente)}$$

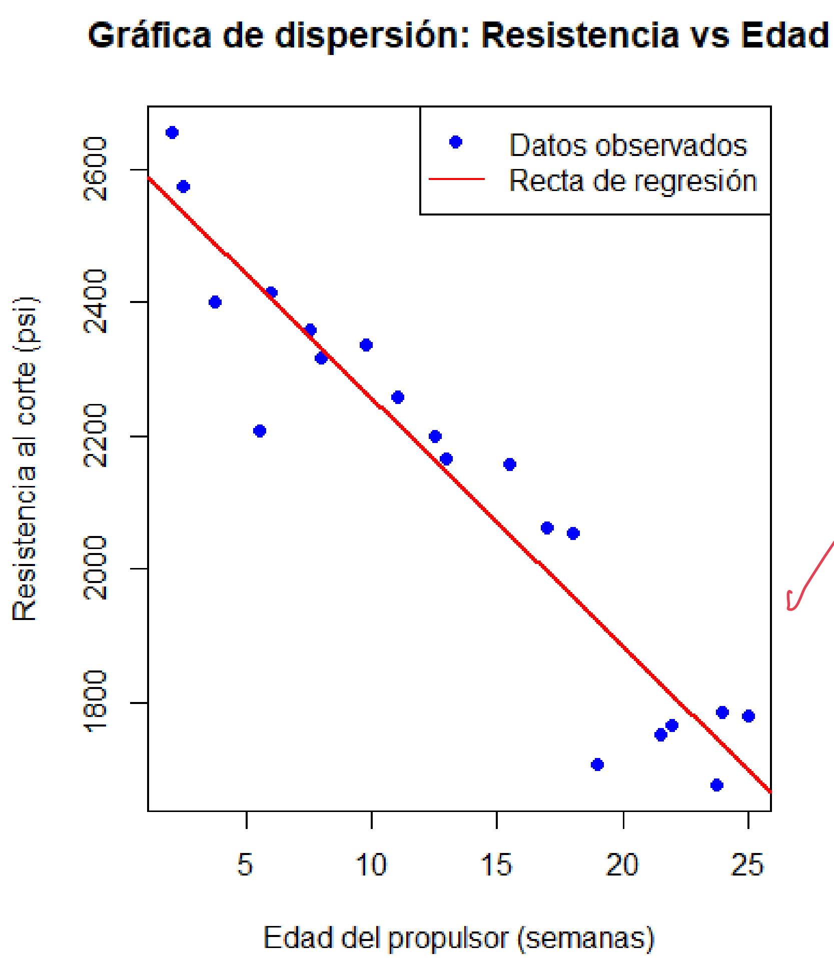
Por tanto, la ecuación de la recta ajustada es:

$$\hat{Y} = 2627.82 - 37.15X$$

Donde:

- Y= Resistencia al corte (psi)
- X= Edad del propulsor (semanas)

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



El ajuste es bueno ya que la mayoría de los puntos están cerca de la línea de regresión, pero se observan ciertos valores atípicos,

5. Efectué la prueba de significancia de la regresión para un nivel α=0.05. Escriba el valor del p − valor. ¿Qué conclusiones puede hacer sobre β1?

Comparación con α = 0.05 y conclusión

- Nivel de significancia: $\alpha = 0.05$.
- Como **p = 1.64e-10 < 0.05**, **rechazamos H_0**.

Conclusión, hay evidencia estadísticamente significativa (al nivel 5%) de que $\beta_1 \neq 0$. La variable Edad tiene un efecto significativo sobre la Resistencia. Como $\hat{\beta}_1 = -37.154$, la relación es negativa a mayor edad, menor 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)?

Usando la fórmula que obtuvimos:

$$\hat{Y} = 2627.82 - 37.15X$$

Ahora sustituimos X =edad en semanas para cada lote:

- Para $X = 5$semanas:

$$\hat{Y} = 2627.82236 + (-37.15359) \times 5 = 2627.82236 - 185.76795 = 2442.05441$$

Estimación: 2442.05 psi (aprox.)

- Para $X = 10$semanas:

$$\hat{Y} = 2627.82236 + (-37.15359) \times 10 = 2627.82236 - 371.53590 = 2256.28646$$

Estimación: 2256.29 psi (aprox.)

- Para $X = 15$semanas:

$$\hat{Y} = 2627.82236 + (-37.15359) \times 15 = 2627.82236 - 557.30385 = 2070.51851$$

Estimación: 2070.52 psi (aprox.)

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

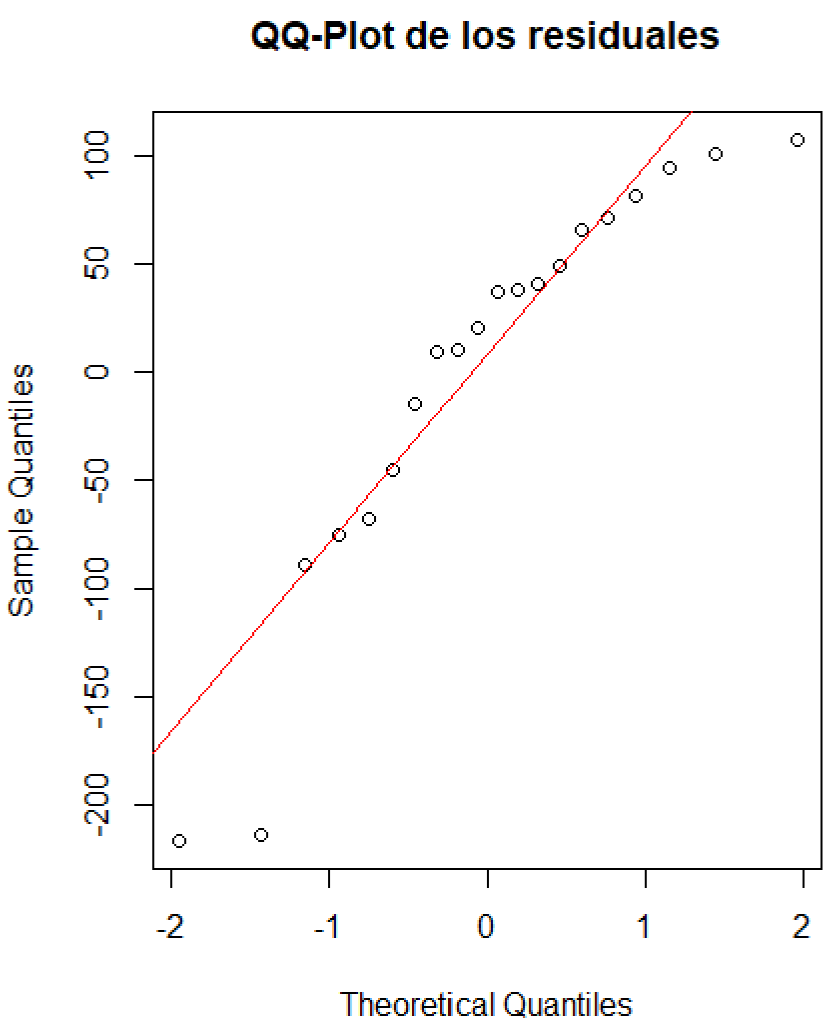
Multiple R-squared: 0.9018

significa que **aproximadamente el 90.18 % de la variabilidad** en la resistencia se explica por la edad del propulsor.

Eso quiere decir que el ajuste es alto

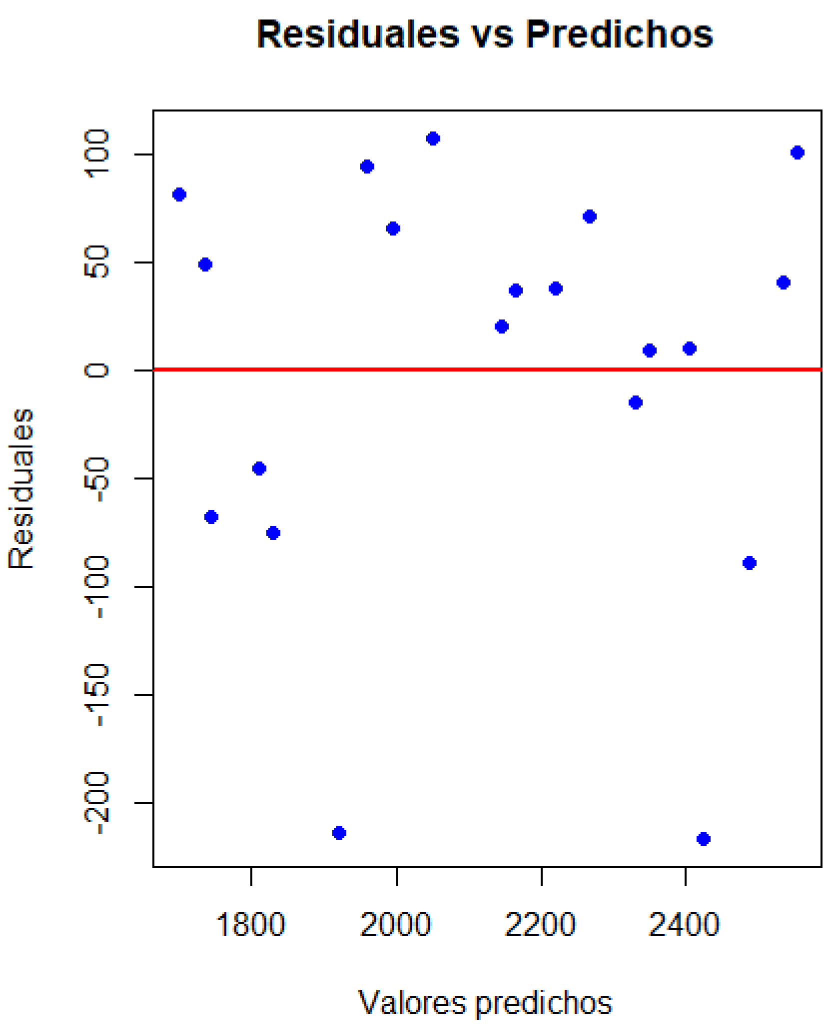
8. VERIFICACION DE SUPUESTOS DEL MODELO. (Obtenga primero los residuales)

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



El QQ-plot de los residuales muestra que la mayoría de los puntos se alinean con la línea de referencia, por lo que el supuesto de normalidad se cumple aproximadamente

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 no oscilan alrededor del eje 0 , además hay más del doble de puntos en la parte superior eso quiere decir que el modelo subestima.

no necesariamente

Varianza constante

La dispersión de los residuales no es completamente uniforme. Se observan algunos puntos con valores residuales extremos, aproximadamente entre -200 y +100, lo que podría indicar una mayor variabilidad en los extremos. Esto sugiere la presencia de heterocedasticidad ligera.

Independencia

No se observa un patrón claro en entre los puntos por lo cual parecen independientes.

9. PUNTOS ATIPICOS E INFLUYENTES.

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

Si tenemos puntos atípicos que oscilan en valores muy altos como +100 y valores muy bajos -200.

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

Estas son las distancias de Cook para cada observación:

1: 0.0373

2: 0.0497

3: 0.0010

4: 0.0161

5: 0.3344

6: 0.2291

7: 0.0270

8: 0.0191

9: 0.0004

10: 0.0047

11: 0.0012

12: 0.0761

13: 0.0889

14: 0.0192

15: 0.0166

16: 0.0387

17: 0.0006

18: 0.0042

19: 0.1317

20: 0.0425

Di>4/n

donde n es el número de observaciones.

- $n = 20$
- Entonces $4/n = 4/20 = 0.2$

Vemos que los puntos 5 y 6 son influyentes.

10. Escriba una conclusión general para este problema.

En base a todos los puntos que obtuvimos podemos concluir que existe una relación entre la edad y la resistencia al corte. La pendiente estimada, $\hat{\beta}_1$ = -37.15, indica que a mayor edad del propulsor, la resistencia tiende a disminuir, confirmando una relación negativa. El p-valor asociado a β_1 ($p \approx 1.64e\text{-}10$) es mucho menor que el nivel de significancia $\alpha = 0.05$, lo que respalda que la variable Edad tiene un efecto significativo sobre la Resistencia. El coeficiente de determinación ($R^2 \approx 0.90$) muestra que aproximadamente el 90 % de la variabilidad en la resistencia se explica por la edad del propulsor, lo que indica un ajuste alto y consistente del modelo. La gráfica de dispersión con la recta de regresión refuerza esta conclusión, ya que la mayoría de los puntos se encuentra cerca de la línea ajustada, aunque se observan ciertos valores atípicos.

En resumen, el modelo es adecuado para describir la relación entre edad y resistencia, la variable Edad tiene un efecto significativo, y el ajuste es alto.