

Estadística y Diseño de Experimentos

Ejercicios 2

13 de Junio, 2025

Nombre del alumnx: Martínez Buenrostro Jorge Rafael

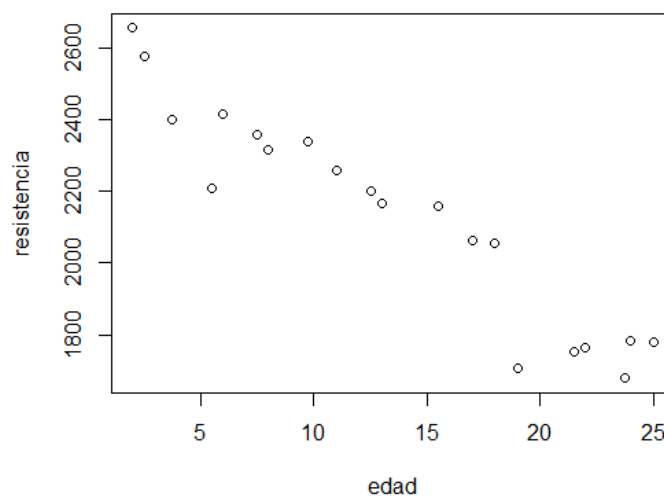
En la fabricación de un motor se deben unir dos tipos de propulsores (tipo 1 y tipo 2). Se sospecha que la resistencia al corte de esta unión está relacionada con la edad (en semanas) del lote de propulsores del tipo 1. En la siguiente tabla se muestra la Resistencia al corte (medida en psi) y la Edad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

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

La variable de respuesta Y es la Resistencia al corte (medida en psi) y la variable regresora X es la Edad (en semanas) del lote del propulsor tipo 1. El modelo de regresión es:

$$Resistencia = \beta_0 + \beta_1 Edad + \epsilon$$

2. Grafique el diagrama de dispersión de los datos.



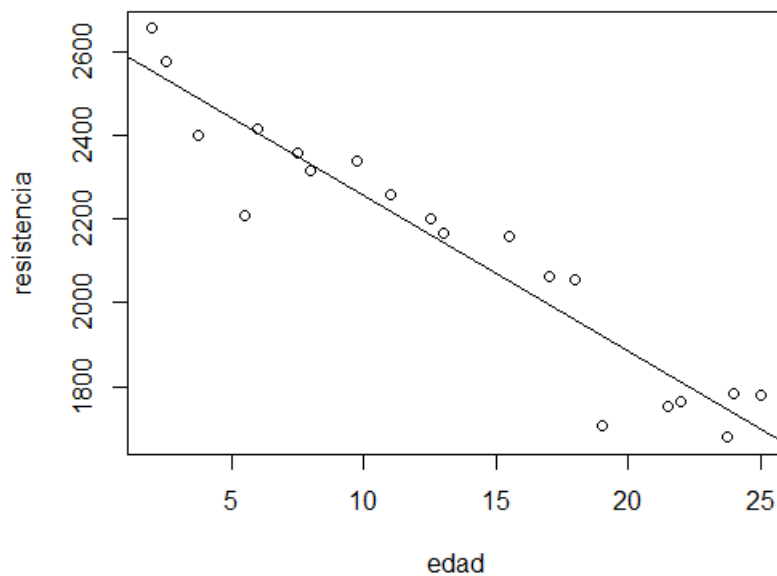
3. Obtenga los estimadores para β_0 y β_1 y escriba la ecuación de la recta ajustada.

```
lm(resistencia~edad)
modelo<-lm(resistencia~edad)
abline(modelo)
summary(modelo)
```

Para obtener los estimadores de β_0 y β_1 , se puede utilizar la función `lm()` y la función `summary()` en R. Los resultados son: $\beta_0 = 2627,82$ y $\beta_1 = -37,154$, y la ecuación de la recta ajustada es:

$$Resistencia = 2627,82 - 37,15 * Edad$$

4. Grafique la recta de la regresión junto con los datos. ¿Qué tan bueno cree que es el ajuste? El ajuste parece ser bueno, ya que la recta de regresión pasa cerca de la mayoría de los puntos del diagrama de dispersión.



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 β_1 ?

La prueba de significancia consiste en comparar el p-valor con α . Si el p-valor es menor que α se rechaza la hipótesis nula. En este caso, el p-valor lo obtenemos de la función `summary()`, y es de $1,64e - 10$. Como $1,64e - 10 < 0,05$, se rechaza la hipótesis nula y se concluye que β_1 es significativamente diferente de cero, lo que indica que la edad del lote del propulsor tipo 1 tiene un efecto significativo en 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)?

```
> predict(modelo, list(edad=c(5)), interval = 'prediction')
      fit      lwr      upr
1 2442.054 2229.021 2655.087
> predict(modelo, list(edad=c(10,15)), interval = 'prediction')
      fit      lwr      upr
1 2256.286 2048.385 2464.188
2 2070.518 1863.382 2277.655
```

La estimación para la resistencia según el modelo de regresión es:

- Para 5 semanas de edad: 2442.054 psi
- Para 10 semanas de edad: 2256.286 psi
- Para 15 semanas de edad: 2070.518 psi

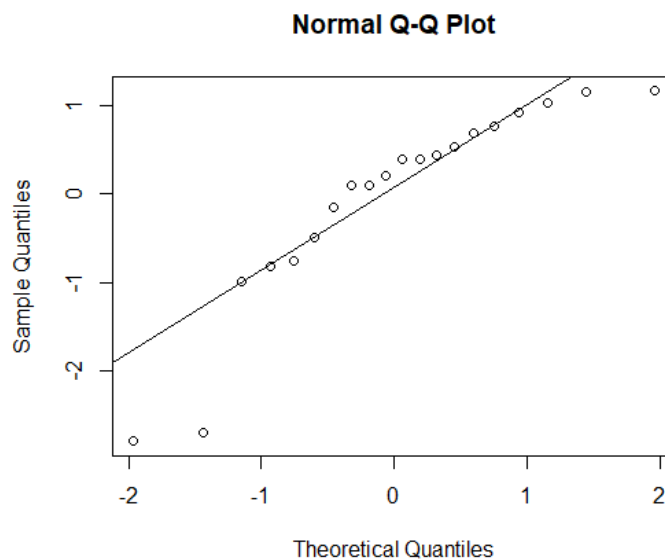
7. Calcule el valor del coeficiente de determinación (Multiple R Squared). Según este coeficiente. ¿qué tan bueno es el ajuste de la regresión?

El valor del coeficiente de determinación (Multiple R Squared) se obtiene de la función `summary()` y es de 0,9018. Esto indica que el 90.18 % de la variabilidad en la resistencia al corte puede ser explicada por la edad del lote del propulsor tipo 1, lo que sugiere que el ajuste de la regresión es bueno.

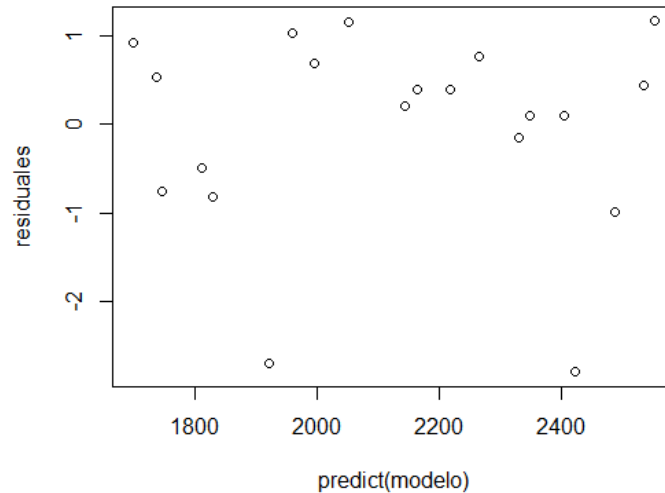
8. VERIFICACIÓN DE SUPUESTOS DEL MODELO. (Obtenga primero los residuales)

Para obtener los residuales se puede utilizar la función `rstudent()` o la función `residuals()` en R. Para este caso, se utilizará la función `rstudent()`.

- a) NORMALIDAD. Grafique los residuales contra los cuantiles de una normal (`qqnorm, qqline`). ¿Se satisface este supuesto? Si, se satisface este supuesto, ya que los puntos se encuentran cerca de la línea recta.



- b) MEDIA CERO, VARIANZA CONSTANTE E INDEPENDENCIA. Grafique los residuales contra los predichos para verificar los tres supuestos (predict, rstudent).



¿Observa alguna anomalía? No, ya que los datos se encuentran distribuidos aleatoriamente alrededor de cero (media igual a 0 y errores independientes) y se pueden delimitar dos bandas paralelas (varianza constante).

9. PUNTOS ATÍPICOS E INFLUYENTES.

- a) Utilizando la gráfica anterior, ¿se observan puntos que puedan considerarse como atípicos (outliers)? Si, hay dos puntos que pueden considerarse como atípicos, ya que se encuentran fuera de las bandas paralelas.
- b) Utilizando la distancia de Cook, verifique si hay puntos influyentes. Para poder calcular la distancia de Cook, se puede utilizar la función `cooks.distance()` en R. Si algún dato tiene una distancia de Cook mayor a 1, se considera un punto influyente.

```
> cooks.distance(modelo)
      1      2      3      4      5
0.0373281981 0.0497291858 0.0010260760 0.0161482719 0.3343768993
      6      7      8      9     10
0.2290842436 0.0270491200 0.0191323748 0.0003959877 0.0047094549
     11     12     13     14     15
0.0012482345 0.0761514881 0.0889892211 0.0192517639 0.0166302585
     16     17     18     19     20
0.0387158541 0.0005955991 0.0041888627 0.1317143774 0.0425721512
> cooks.distance(modelo)>1
      1      2      3      4      5      6      7      8      9     10     11
FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
     12     13     14     15     16     17     18     19     20
FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

10. Escriba una conclusión general para este problema.

Los resultados del análisis de regresión indican que existe una relación significativa entre la edad del lote del propulsor tipo 1 y la resistencia al corte. El modelo de regresión lineal ajustado sugiere que a medida que aumenta la edad del lote, la resistencia al corte disminuye. El coeficiente de determinación indica que el modelo explica una gran parte de la variabilidad en la resistencia al corte. Los supuestos del modelo parecen cumplirse adecuadamente, y no se identificaron puntos influyentes significativos. En general, se puede concluir que la edad del lote del propulsor tipo 1 es un factor importante a considerar en la fabricación de motores para asegurar una resistencia adecuada al corte.

Código en R

```
> #Carga de 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)
>
>
> # 1. Identificar variables
> # Variable de respuesta Y: Resistencia
> # Variables regresora: Edad
> # Modelo:  $Y = B_0 + B_1X + d$ 
>
>
> # 2. Diagrama de dispersion de los datos
> plot(edad,resistencia)
>
> # 3. Estimadores para  $B_0$  y  $B_1$ 
> lm(resistencia~edad)

Call:
lm(formula = resistencia ~ edad)

Coefficients:
(Intercept)      edad
    2627.82     -37.15

>
> modelo<-lm(resistencia~edad)
> abline(modelo)
> 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 ***
---
>
> # Ecuacion de la recta ajustada
> cat("Resistencia =", round(beta0, 2), "+", round(beta1, 2), "* Edad\n")
Error: objeto 'beta0' no encontrado
```

```

> cat("Resistencia =", round(coef(modelo)[1], 2), "+", round(coef(modelo)[2],
  2), "* Edad\n")
Resistencia = 2627.82 + -37.15 * Edad
>
> # 4. Grafica de la recta de regresion con datos
> plot(edad,resistencia)
> abline(modelo)
>
>
> # 5. Prueba de significancia, p-valor<alpha
> # p-valor = 1.64e-10
> # alpha = 0.05
>
>
> # 6. Estimacion para lotes tipo 1 con 5, 10 y 15
> predict(modelo, list(edad=c(5)), interval = 'prediction')
      fit      lwr      upr
1 2442.054 2229.021 2655.087
> predict(modelo, list(edad=c(10,15)), interval = 'prediction')
      fit      lwr      upr
1 2256.286 2048.385 2464.188
2 2070.518 1863.382 2277.655
>
>
> # 7. Coeficiente de determinacion
> # Multiple R-squared: 0.9018
>
> # 8. Verificacion de supuestos del modelo
> residuales<-residuals(modelo)
> residuales<-rstudent(modelo)
> qqnorm(residuales)
> plot(predict(modelo),residuales)
>
>
> # 9. Puntos atipicos e influyentes
> cooks.distance(modelo)
      1      2      3      4      5
0.0373281981 0.0497291858 0.0010260760 0.0161482719 0.3343768993
      6      7      8      9     10
0.2290842436 0.0270491200 0.0191323748 0.0003959877 0.0047094549
     11     12     13     14     15
0.0012482345 0.0761514881 0.0889892211 0.0192517639 0.0166302585
     16     17     18     19     20
0.0387158541 0.0005955991 0.0041888627 0.1317143774 0.0425721512
> cooks.distance(modelo)>1
      1      2      3      4      5      6      7      8      9     10     11     12
FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
     13     14     15     16     17     18     19     20
FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
> outliers<-which(abs(residuales) > 2)

```

```
> if(length(outliers) > 0) {  
+   print(outliers)  
+ } else {  
+   cat("No se detectaron outliers significativos\n")  
+ }  
5 6  
5 6  
>
```