



ESTADISTICA Y DISEÑO DE EXPERIMENTOS

Ejercicios 2

Juan Diego Ayala Velázquez

Alexis Noe Cruz Ramírez

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

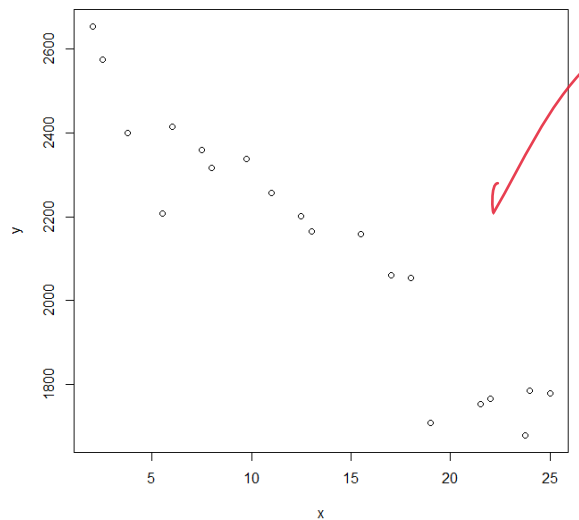
Cuadro 1: Resistencia obtenida según la edad del propulsor tipo 1.

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

- La variable de respuesta Y es la columna Resistencia
- La variable regresora X es la columna de Edad
- $Y = \beta_0 + \beta_1 X$

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

```
"y <- c(2158.70, 1678.15, 2316.00, 2061.30, 2207.50, 1708.30, 1784.70, 2575.00, 2357.90
2256.70"
> y <- 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)
> y
[1] 2158.70 1678.15 2316.00 2061.30 2207.50 1708.30 1784.70 2575.00 2357.90
[10] 2256.70 2165.20 2399.55 1779.80 2336.75 1765.30 2053.50 2414.40 2200.50
[19] 2654.20 1753.70
> x <- 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)
> x
[1] 15.50 23.75 8.00 17.00 5.50 19.00 24.00 2.50 7.50 11.00 13.00 3.75
[13] 25.00 9.75 22.00 18.00 6.00 12.50 2.00 21.50
> plot(x, y)
```



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

```
> RL <- lm(y ~ x)
> RL$coefficients
(Intercept)          x
 2627.82236    -37.15359
> summary(RL)

Call:
lm(formula = y ~ x)

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 ***
x           -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
```

Ecuación de la recta ajustada:

$$Y = 2627.822 - 37.154X$$

Coefficiente $\beta_1 = -37.154$ nos dice que, por cada semana adicional de edad, la resistencia p. disminuye en aproximadamente 37.154 psi

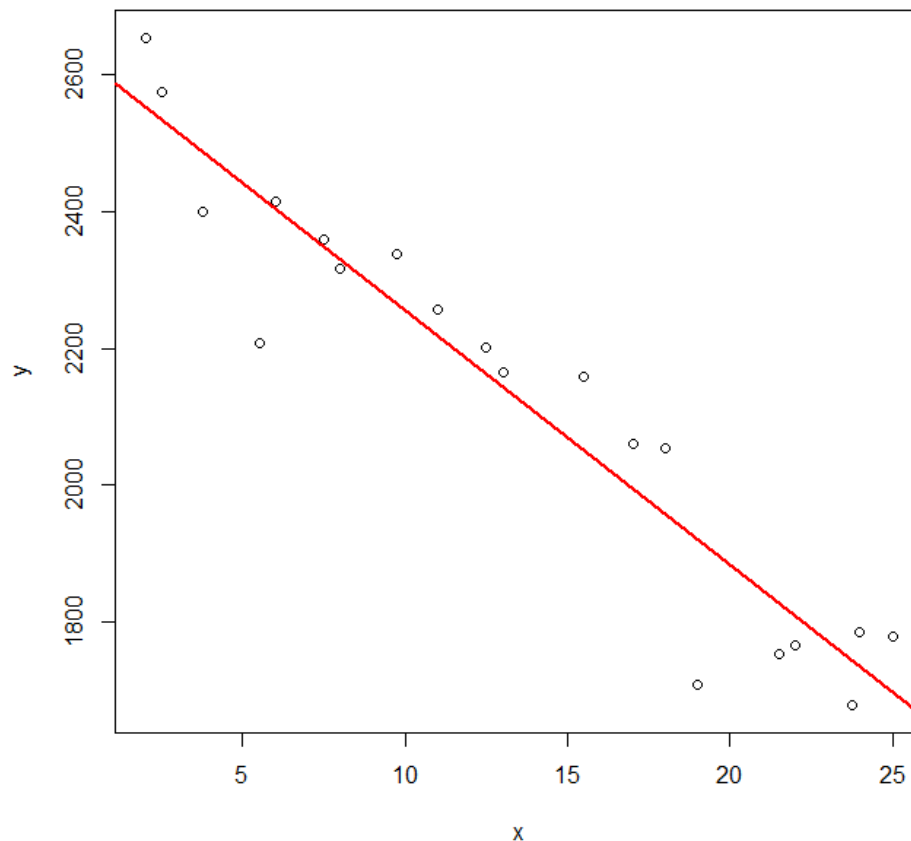
Coefficiente $\beta_0 = 2627.822$ representa la resistencia estimada para un propulsor recién fabricado

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

$$R^2 = 0.9018$$

El ajuste del modelo creo que es muy bueno, yo diría que es excelente, con este dato aproximadamente el 90.18% de la variabilidad en la resistencia se explica por la edad del propulsor y la gráfica nos está mostrando que todos los puntos se encuentran muy cerca de la recta

```
> abline(RL, col="red", lwd=2)
> |
```



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

$\alpha=0.05$.

Entonces p – valor = $1.643344e-10$

Se rechaza H_0 , la regresión fue estadísticamente significativa como para afirmar que esa relación entre la edad del propulsor tipo 1 y la resistencia al corte es significativa siendo $p < 0.5$. Entonces la variable edad es un predictor estadísticamente válido de 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)?

```
> pred <- predict(modelo, newdata = nuevas_edades, interval = "prediction", level = 0.95)
> pred
```

	fit	lwr	upr
1	2442.054	2229.021	2655.087
2	2256.286	2048.385	2464.188
3	2070.518	1863.382	2277.655

1) Para Edad = 5,

2) Para Edad = 10,

3) Para Edad = 15

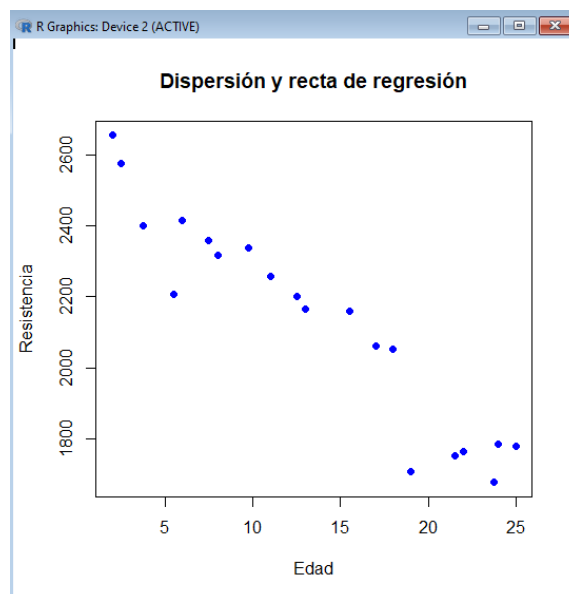
7. Calcule el valor del coeficiente de determinación (Multiple R Squared). Según este coeficiente. ¿qué tan bueno es el ajuste de la regresión?

```
> cat("R² =", summary(modelo)$r.squared, "\n")
R² = 0.9018414
> cat("R² ajustado =", summary(modelo)$adj.r.squared, "\n")
R² ajustado = 0.8963882
```

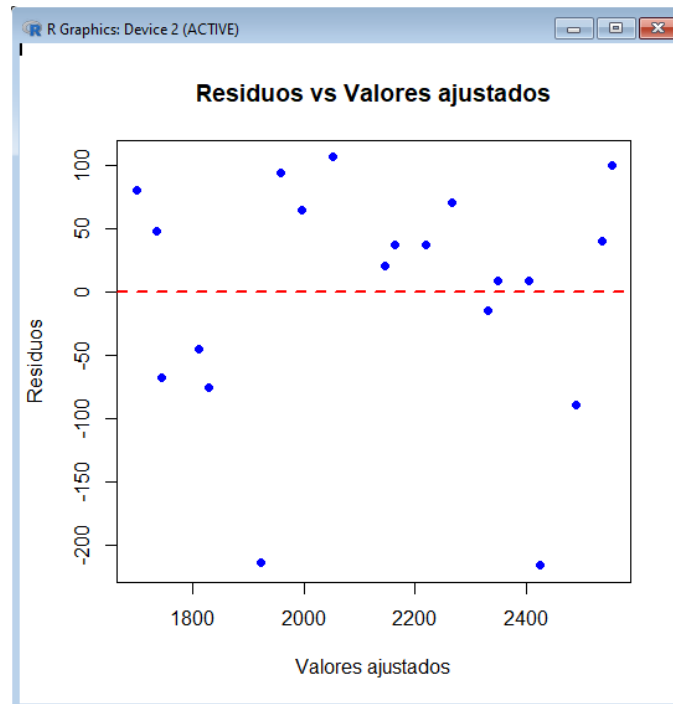
8. VERIFICACION DE SUPUESTOS DEL MODELO. (Obtenga primero los residuales)

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

Falta la verificación de los supuestos de normalidad



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



A simple vista, los datos parecen cumplir con las condiciones necesarias: tienen una forma más o menos normal, su promedio está cerca de cero y la variación no cambia mucho. No se ve ninguna señal clara en los gráficos que indique que algo esté mal.

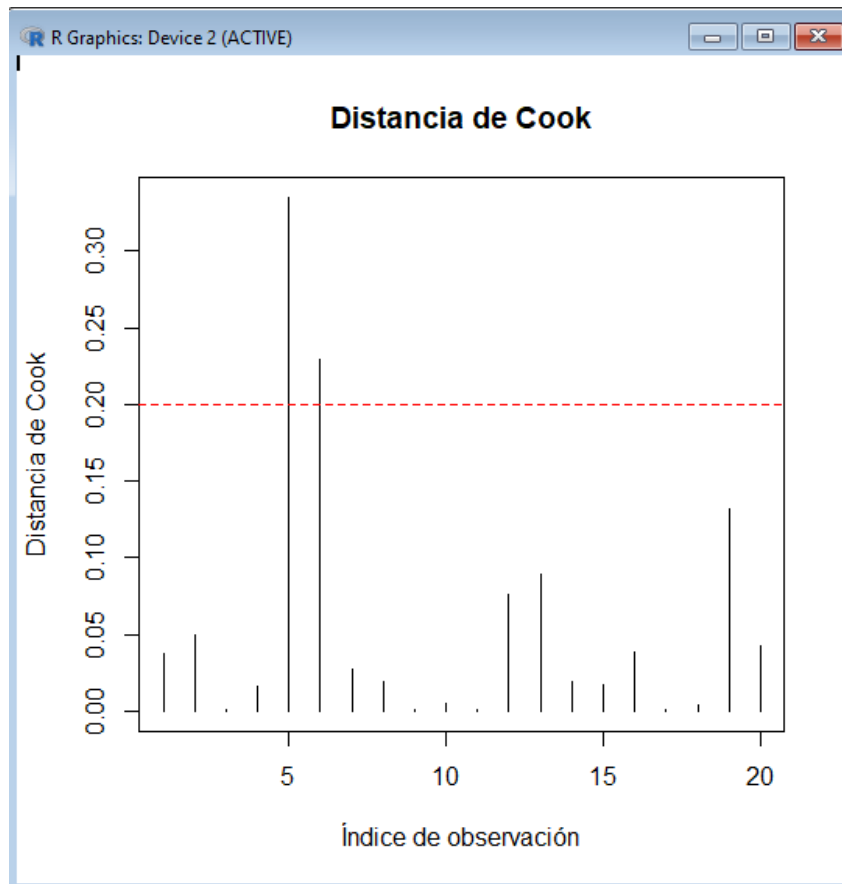
9. PUNTOS ATÍPICOS E INFLUYENTES.

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

En la gráfica que compara los errores con los valores estimados, no se ven errores que sean mucho más grandes que los demás. Además, en la tabla, los errores ajustados tienen valores normales, y ninguno es tan alto como para preocuparse.

Por lo tanto, no hay outliers claros según el criterio visual.

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



No se encontró ningún dato que tenga una influencia muy fuerte en el modelo. Según una regla práctica, si el valor de Cook es mayor que 0.2 (porque hay 20 datos), podría indicar que ese dato está afectando demasiado los resultados. Pero en este caso, todos los valores están dentro de lo normal.

Por lo tanto, no hay puntos influyentes evidentes según la Distancia de Cook.

10. Escriba una conclusión general para este problema.

Hay una relación clara y negativa entre la edad del lote del propulsor tipo 1 y su resistencia al corte. Esto significa que, a medida que el propulsor envejece, su resistencia se va reduciendo. Además, el modelo RLS logra explicar casi el 90% de los cambios en la resistencia, lo que indica que el modelo se ajusta muy bien a los datos.

Los supuestos clásicos (normalidad aproximada de errores, media cero, homocedasticidad e independendencia) no muestran violaciones fuertes a partir de las pruebas gráficas