# Clase Práctica 6
# Regresión Lineal Simple

Equipo 1
Dalianys Pérez Perera
Dayany Alfaro González
Gilberto González Rodríguez
Grupo C-411

# Índice

# 1. Introducción

La regresión lineal simple consiste en generar un modelo de regresión(ecuación de la recta) que permita explicar la relación lineal que existe entre dos variables. Una recta de regresión puede emplearse para diferentes propósitos y dependiendo de ellos es necesario satisfacer distintas condiciones.

En caso de querer medir la relación entre dos variables, la recta de regresión lo va a indicar de forma directa. Sin embargo, en caso de querer predecir el valor de una variable en función de la otra, no solo se necesita calcular la recta, sino que además hay que asegurar que el modelo sea bueno. Es importante tener en cuenta que que las predicciones deben, a priori, limitarse al rango de valores dentro del que se encuentran las observaciones con las que se ha generado el modelo. Esto es importante puesto que se solo en esta región se tiene certeza de que se cumplen las condiciones para que el modelo sea válido.

En el presente trabajo se analizan tres diferentes casos para determinar la relación existente entre sus variables, además de la estimación de valores a partir del modelo obtenido y la comprobación de los supuestos para dicho modelo.

## 2. Ejercicio 1

Ajuste los siguientes datos a una recta por mínimos cuadrados:

| $x$ | -1 | 0 | 3 | 7 |
|---|---|---|---|---|
| $f(x)$ | 2 | 0 | 4 | 7 |

Cuadro 1: Datos iniciales

**Test de correlación de Pearson:**
$r = 0{,}924037$
intervalo de confianza(95 %) $[-0{,}33104226, \quad 0{,}9984344]$
$p - valor$ : 0,07596

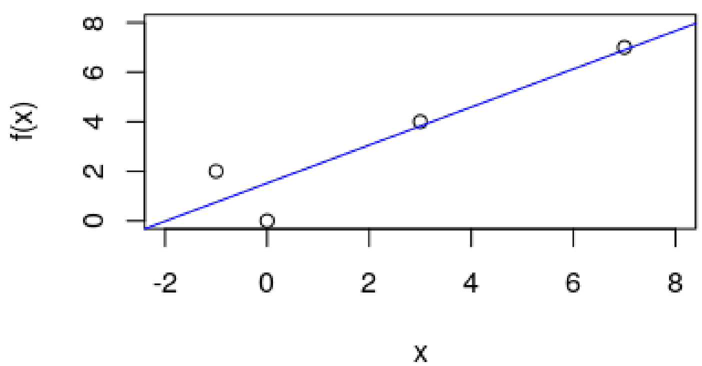


Figura 1: Diagrama de dispersión

El primer paso antes de generar un modelo de regresión lineal es representar los datos para poder intuir si existe una relación y cuantificar dicha relación mediante un coeficiente de correlación. La figura 1 y el test de correlación muestran una relación lineal de intensidad considerable( $r = 0{,}924037$) y significancia($p - valor$ : 0,07596).

**Ecuación del modelo:** $y = 1{,}5226x + 0{,}7677$

**Estimación**
Para estimar el valor de $x = 1$ se utilizó la función predict de R y se obtuvo:
$f(1) = 2{,}290323$
intervalo de confianza(95 %) $[-0{,}951074, \quad 5{,}531719]$

## 2.1. Diagnóstico del modelo

- **Linealidad de los errores .** Para analizar este supuesto se aplicó la prueba Durbin-Watson obteniendo como resultado:

$$DW = 2{,}697$$
$$p - valor = 0{,}5136$$
$$\alpha = 0{,}05$$

Interpretación: Como $p - valor > \alpha$ no se puede rechazar $H_0$ por lo que se puede decir que los errores son independientes.

- **Valor esperado del error aleatorio igual a cero.**

| Min | 1er Qu | Mediana | Media | 3er Qu | Max |
|---|---|---|---|---|---|
| -1.5226 | -0.3032 | 0.1387 | 0.0000 | 0.4419 | 1.2452 |

Cuadro 2: Residuos

Se comprueba con el Cuadro 2 que el valor esperado del error aleatorio es cero, además en la figura 2 donde en el eje $x$ se encuentran los valores ajustados y en el eje de las ordenadas sus respectivos errores, se muestra que estos últimos están cercanos a cero.

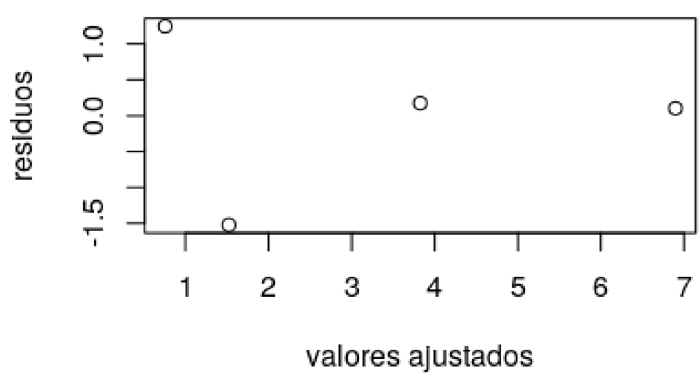


Figura 2: Gráfico de residuos vs datos

- **Homocedasticidad.**

Para analizar este supuesto se aplicó la prueba Breusch-Pagan obteniendo como resultado:

$$BP = 2{,}7038$$
$$p - valor = 0{,}1001$$
$$\alpha = 0{,}05$$

Interpretación: Como $p - valor > \alpha$ no se puede rechazar $H_0$ por lo que se puede decir que se cumple la Homocedasticidad.

- **Los errores además de ser independientes son idénticamente distribuidos y siguen distribución normal con media cero y varianza constante.**

  Para analizar este supuesto primeramente se recurrió al test de contraste de normalidad Shapiro-Wilk dando como resultado:

$$W = 0{,}93698$$
$$p - valor = 0{,}636$$
$$\alpha = 0{,}05$$

  Interpretación: Como $p - valor > \alpha$ no se puede rechazar la hipótesis nula por tanto los errores siguen una distribución normal.

  También se analizó el histograma de frecuencias y los cuantiles normales, ver figuras 3 y 4

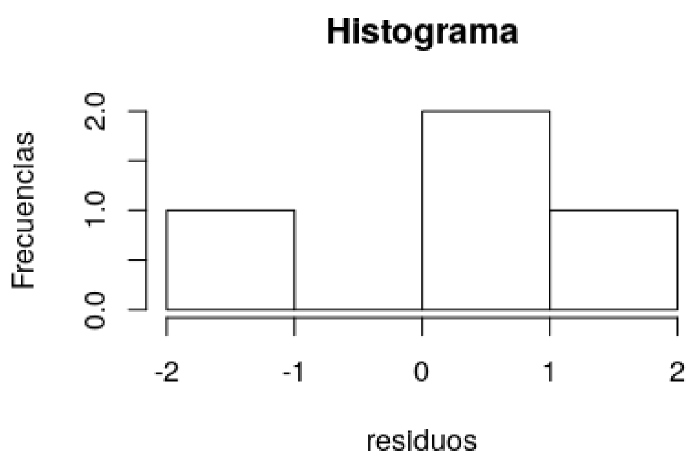


Figura 3: Histograma de los residuos

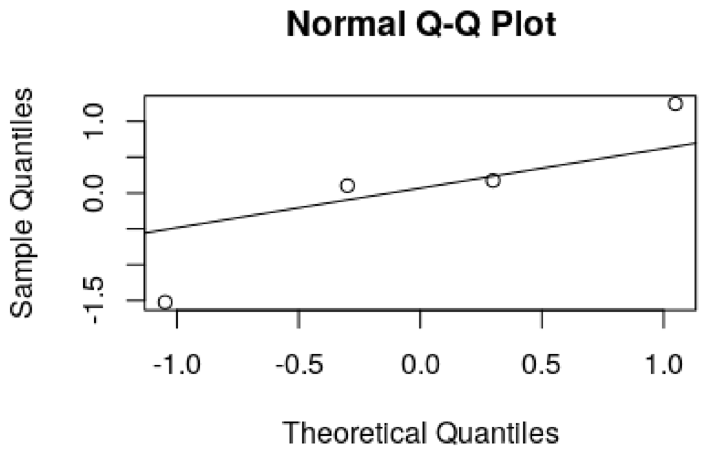


Figura 4: QQ-plot de los residuos

A pesar de ser pocas observaciones, tanto la representación gráfica como el contraste de hipótesis confirman la distribución normal de los residuos con $\mu = 0$ y varianza constante.

# 3. Ejercicio 2

Ajuste los siguientes datos a una recta por mínimos cuadrados:

| $x$ | -3 | -1 | 1 | 3 | 5 | 7 |
|---|---|---|---|---|---|---|
| $f(x)$ | 14 | 4 | 2 | 8 | 22 | 44 |

Cuadro 3: Datos iniciales

**Test de correlación de Pearson:**
$r = 0{,}7164552$
intervalo de confianza(95 %) $[-0{,}2272261, \quad 0{,}9662140]$
$p - valor : 0{,}1092$

**Ecuación del modelo:** $y = 3x + 9{,}667$

**Estimación**

| $x$ | $f(x)$ | intervalo de confianza |
|---|---|---|
| 0 | 9.666667 | $[-6{,}384505, \quad 25{,}71784]$ |
| 2 | 15.666667 | $[1{,}815325, \quad 29{,}51801]$ |

Cuadro 4: Valores estimados

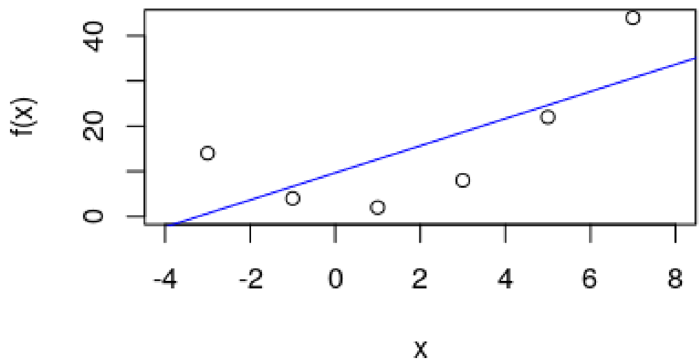


Figura 5: Diagrama de dispersión

## 3.1. Diagnóstico del modelo

- **Linealidad de los errores .** Para analizar este supuesto se aplicó la prueba Durbin-Watson obteniendo como resultado:

$$DW = 1{,}0714$$
$$p - valor = 0{,}003544$$
$$\alpha = 0{,}05$$

Interpretación: Como $p - valor < \alpha$ se rechaza $H_0$ por lo que se puede decir que los errores están correlacionados, no son independientes.

- **Valor esperado del error aleatorio igual a cero.**

| Min | 1er Qu | Mediana | Media | 3er Qu | Max |
|---|---|---|---|---|---|
| -10.667 | -8.667 | -2.667 | 0.000 | 9.333 | 13.333 |

Cuadro 5: Residuos

Se comprueba en el Cuadro 5 que el valor esperado del error aleatorio es cero. Mientras que en la figura 6 a continuación, se observa que algunos de los residuos están distribuidos más alejados de cero que otros, llegando a alcanzar 10 unidades de diferencia.

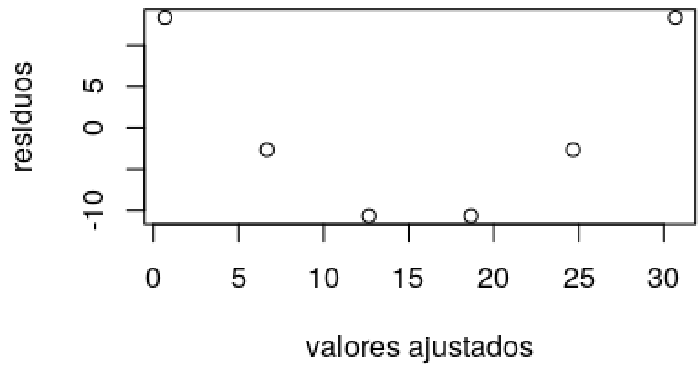


Figura 6: Gráfico de residuos vs valores ajustados

- **Homocedasticidad.** Para analizar este supuesto se aplicó la prueba Breusch-Pagan obteniendo como resultado:

$$BP = 4{,}8137e - 31$$
$$p - valor = 1$$
$$\alpha = 0{,}05$$

  Interpretación: Como $p - valor > \alpha$ no se puede rechazar $H_0$ por lo que se puede decir que se cumple la Homocedasticidad.

- **Los errores además de ser independientes son idénticamente distribuidos y siguen distribución normal con media cero y varianza constante.**

  Para analizar este supuesto primeramente se recurrió al test de contraste de normalidad Shapiro-Wilk dando como resultado:

$$W = 0{,}82272$$
$$p - valor = 0{,}09315$$
$$\alpha = 0{,}05$$

  Interpretación: Como $p - valor > \alpha$ no se puede rechazar la hipótesis nula por tanto los errores siguen una distribución normal.

  En este caso el histograma de frecuencias no muestra ser una distribución normal, no obstante la prueba de hipótesis realizada y la figura 8 sí lo confirman.

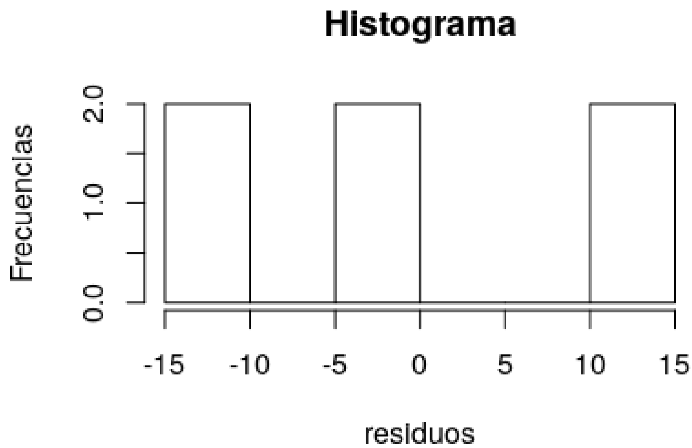


Figura 7: Histograma de los residuos

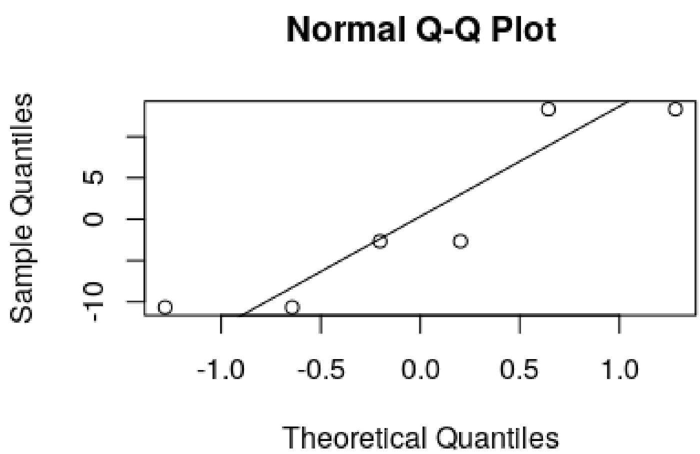


Figura 8: QQ-plot de los residuos

Dado que el modelo no cumple con la linealidad del error aleatorio se concluye que no es un modelo bueno para realizar predicciones al no cumplir con uno de los supuestos, además el coeficiente de Pearson $r = 0{,}7164552$ indica no ser tan fuerte la relación lineal.

# 4. Ejercicio 3

Se tiene un muelle sometido a tracción, se ha cargado el muelle con diferentes pesos y se han anotado los alargamientos:

| $CargasSucesivas(gramos)$ | $Alargamientos$ |
|---|---|
| 200 | 60 |
| 400 | 120 |
| 500 | 150 |
| 700 | 210 |
| 900 | 260 |
| 1000 | 290 |

Cuadro 6: Datos iniciales

**Test de correlación de Pearson:**
$r = 0{,}9994979$
intervalo de confianza(95 %) $[0{,}9951835, \quad 0{,}9999478]$
$p - valor = 3{,}781e - 07$

**Ecuación del modelo:** $y = 0{,}28576x + 5{,}44484$

**Estimación**

| $x$ | $f(x)$ | intervalo de confianza |
|---|---|---|
| 0 | 5.44484 | $[-3{,}069108, \quad 13{,}95879]$ |
| 2 | 6.01637 | $[-2{,}474674, \quad 14{,}50741]$ |

Cuadro 7: Valores estimados

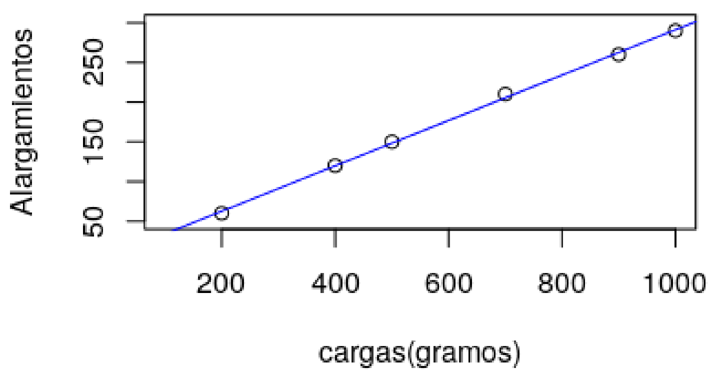


Figura 9: Diagrama de dispersión

La figura 9 y el test de correlación muestran una relación lineal de intensidad considerable( $r = 0{,}9994979$) y significancia($p - valor = 3{,}781e - 07$).

## 4.1. Diagnóstico del modelo

- **Linealidad de los errores .** Para analizar este supuesto se aplicó la prueba Durbin-Watson obteniendo como resultado:

$$DW = 1{,}8585$$
$$p - valor = 0{,}1827$$
$$\alpha = 0{,}05$$

  Interpretación: Como $p - valor > \alpha$ no se puede rechazar $H_0$ por lo que se puede decir que los errores son independientes.

- **Valor esperado del error aleatorio igual a cero.**

| Min | 1er Qu | Mediana | Media | 3er Qu | Max |
|---|---|---|---|---|---|
| -2.6335 | -2.2509 | -0.4804 | 0.0000 | 1.3167 | 4.5196 |

Cuadro 8: Residuos

  Se comprueba con el Cuadro 8 que el valor esperado del error aleatorio es cero, mientras que en la figura 10 se observa que los residuos estn cercanos a cero en comparación con los valores ajustados que son grandes.

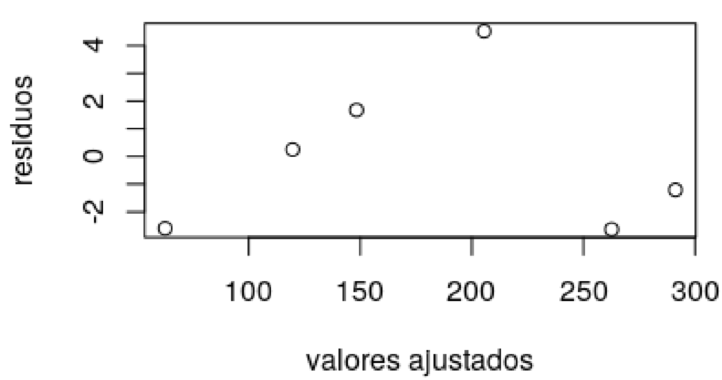


Figura 10: Gráfico de residuos vs datos

- **Homocedasticidad.** Para analizar este supuesto se aplicó la prueba Breusch-Pagan obteniendo como resultado:

$$BP = 0{,}054052$$
$$p - valor = 0{,}8162$$
$$\alpha = 0{,}05$$

Interpretación: Como $p-valor > \alpha$ no se puede rechazar $H_0$ por lo que se puede decir que se cumple la Homocedasticidad.

- **Los errores además de ser independientes son idénticamente distribuidos y siguen distribución normal con media cero y varianza constante.**

  Para analizar este supuesto primeramente se recurrió al test de contraste de normalidad Shapiro-Wilk dando como resultado:

$$W = 0{,}91393$$
$$p - valor = 0{,}4627$$
$$\alpha = 0{,}05$$

  Interpretación: Como $p - valor > \alpha$ no se puede rechazar la hipótesis nula por tanto los errores siguen una distribución normal.

  También se analizó el histograma de frecuencias y los cuantiles normales, ver figuras 11 y 12.

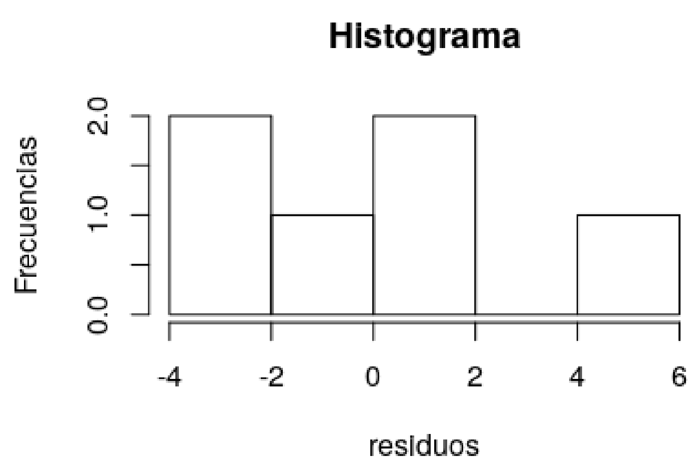


Figura 11: Histograma de los residuos

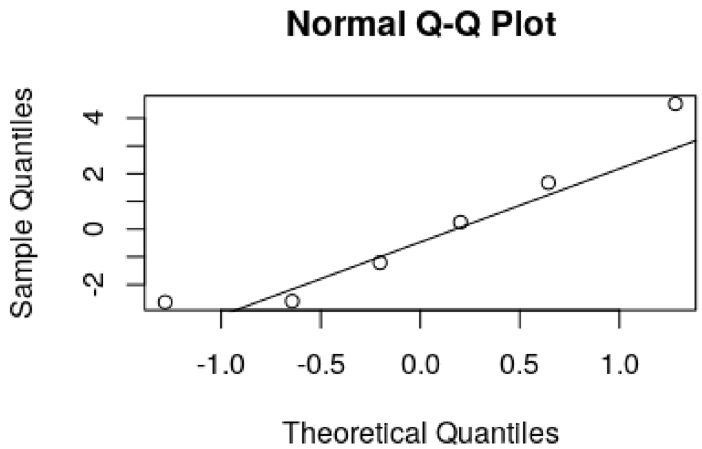


Figura 12: QQ-plot de los residuos

A pesar de ser pocas observaciones, tanto la representación gráfica como el contraste de hipótesis confirman la distribución normal de los residuos con $\mu = 0$ y varianza constante.