REGRESIÓN LINEAL

1. ¿Qué es la regresión?

La regresión es una técnica estadística cuyo objetivo es estudiar y cuantificar la relación entre variables. Permite responder preguntas como:

- ¿Si se conoce X, es posible estimar Y?
- ¿Existe relación entre dos variables?
- ¿Cómo cambia Y cuando cambia X?

La regresión se utiliza para explicar y predecir fenómenos. Algunos ejemplos:

- Relación entre inversión en publicidad y ventas.
- Relación entre tamaño de una vivienda y su precio.
- Relación entre edad y presión arterial.

En toda regresión existen dos variables:

- **Variable dependiente (Y):** la que se desea predecir.
- **Variable independiente (X):** la que se usa para realizar la predicción.

2. ¿Qué es una función lineal?

Una función lineal describe una relación proporcional mediante la expresión:

$$y = mx + b$$

Donde:

- **m** es la pendiente, que indica cuánto cambia 'y' cuando 'x' aumenta una unidad.
- **b** es el intercepto, que representa el valor de 'y' cuando x = 0.

Una función lineal es una ecuación predefinida, no proviene de datos empíricos.

3. Modelo lineal con término de error

En estadística se considera que la relación entre X y Y no es perfecta. Por ello, el modelo teórico incorpora un término que recoge toda la variabilidad no explicada por X:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Donde:

- β_0 es el verdadero intercepto de la población.
- β_1 es la verdadera pendiente de la población.
- ε es el término de error, que recoge:
  - factores no medidos,
  - variabilidad aleatoria,
  - ruido o imprecisiones en la medición.

Este modelo expresa que Y depende de X y además de múltiples factores no observados representados por ε.

4. Recta de regresión estimada

Dado que β_0 y β_1 son desconocidos, se estiman a partir de datos. La recta estimada se expresa como:

$$y\hat{} = mx + b$$

Donde:

- **y^** es el valor estimado o predicho.
- **m** es la estimación de la pendiente β_1.
- **b** es la estimación del intercepto β_0.

Los residuos se definen como:

$$e_i = y_i - y\hat{}_i$$

Y representan el error entre los valores observados y los estimados.

5. Criterio de ajuste: Mínimos Cuadrados Ordinarios (OLS)

La regresión lineal elige la recta que minimiza la suma total de los errores al cuadrado:

$$\sum(y_i - y\hat{}_i)^2$$

Se eleva el error al cuadrado para:

- Evitar que errores positivos y negativos se cancelen.
- Penalizar más los errores grandes.

Este método se conoce como Mínimos Cuadrados Ordinarios (OLS).

6. Ejemplo detallado: cálculo de la pendiente

Se utiliza el siguiente conjunto de datos:

X	Y
1	2
2	3
3	5

La ecuación estimada tendrá la forma: *y^ = mx + b*

Paso 1: Calcular promedios

Promedio X = (1+2+3) / 3 = 2

Promedio Y = (2+3+5) / 3 = 3.33

Paso 2: Calcular la pendiente m

La pendiente se obtiene mediante la siguiente fórmula:

$$\mathbf{m} = \frac{\sum (x_i - x\blacksquare)(y_i - \blacksquare)}{\sum (x_i - x\blacksquare)^2}$$

a) Desviaciones respecto a los promedios

X	Y	x - prom(x)	y - prom(y)
1	2	-1	-1.33
2	3	0	-0.33
3	5	1	1.67

b) Productos de desviaciones (Numerador)

X	Y	Producto
1	2	1.33
2	3	0
3	5	1.67
	SUMA:	**3**

Suma del numerador = 3

c) Cuadrados de las desviaciones (Denominador)

X	x - prom(x)	Cuadrado
1	-1	1
2	0	0
3	1	1
	SUMA:	**2**

Suma del denominador = 2

d) Pendiente final

$$m = 3 / 2 = 1.5$$

Paso 3: Calcular el intercepto b

b = Promedio Y - (m * Promedio X)

b = 3.33 - (1.5 * 2)

b = 0.33

7. Recta de regresión final

$$y\hat{} = 1.5x + 0.33$$

Interpretación:

- Por cada aumento de 1 unidad en X, Y aumenta en promedio 1.5 unidades.
- Cuando X = 0, el valor estimado de Y es aproximadamente 0.33.

8. Calidad del ajuste: coeficiente R^2

El coeficiente de determinación R^2:

- Toma valores entre 0 y 1.
- Indica qué porcentaje de la variación de Y es explicado por el modelo.

Ejemplo: $R^2 = 0.85$ significa que el 85% de la variación observada en Y puede explicarse mediante X. Cuanto más se aproxime a 1, mejor es el ajuste del modelo.