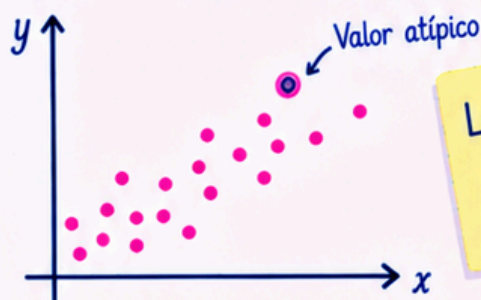


1 Datos bidimensionales

Tenemos pares de datos (x_i, y_i) con $i = 1, \dots, N$. Se representa en un diagrama de dispersión (nube de puntos) y se observa la tendencia, la forma y la posible existencia de valores atípicos.



La correlación no prueba causalidad.

2 Medias y covarianza

Medias de las variables:

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} \quad \bar{y} = \frac{\sum_{i=1}^N y_i}{N}$$

Covarianza:

$$\sigma_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{N} = \frac{\sum_{i=1}^N x_i y_i}{N} - \bar{x} \bar{y}$$

Si los datos tienen frecuencias, se ponderan las sumas por sus frecuencias.

3 Correlación lineal

El coeficiente de correlación lineal de Pearson se define como

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \text{ si } \sigma_x, \sigma_y > 0.$$

- $-1 \leq r \leq 1$.
- El signo de r indica el sentido de la relación lineal.
- El valor absoluto $|r|$ indica la intensidad de la relación lineal.
- $r \approx 0$ no descarta la existencia de una relación no lineal.

r mide la fuerza y el sentido de la relación lineal entre las variables.

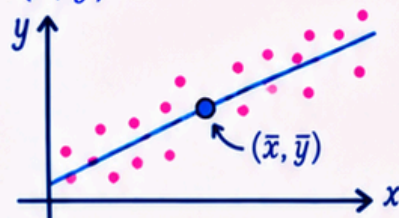
4 Regresión de y sobre x

La recta de regresión de y sobre x es

$$\hat{y} = \bar{y} + b_{yx}(x - \bar{x})$$

donde $b_{yx} = \frac{\sigma_{xy}}{\sigma_x^2}$

- Permite estimar el valor de y a partir de un valor de x .
- La recta pasa por el punto $(\bar{x}, \bar{y})$.



5 Regresión de x sobre y

La recta de regresión de x sobre y es

$$\hat{x} = \bar{x} + b_{xy}(y - \bar{y})$$

donde $b_{xy} = \frac{\sigma_{xy}}{\sigma_y^2}$

- Permite estimar el valor de x a partir de un valor de y .
- No se obtiene simplemente despejando la otra recta.

$$b_{yx} \cdot b_{xy} = r^2.$$

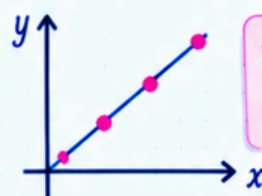
Analiza, interpreta y decide con criterio

Los datos te cuentan historias, tú les das sentido

6 Predecir con sentido

Ejemplo: datos (1, 2), (2, 4), (3, 6).

x	y
1	2
2	4
3	6



$$r = 1$$

$$\hat{y} = 2x$$

- Interpolar es estimar un valor dentro del intervalo de valores de x observado.
- Extrapolar es estimar un valor fuera del intervalo observado y es más incierto.
- No conviertas una estimación en certeza.