Correlación de cantidades físicas

Basado en el capítulo '10: Correlation of physical quantities' del libro The Uncertainty in Physical Measurements de Paolo Fornasini

27 de febrero de 2025

Brenda **Alderete**, Gastón **Barcala Berengard**, Sol **Berns**

¿Por qué nos interesa estudiar la correlación?

- Correlación: relación funcional entre dos (o más) magnitudes.
- Permiten inducir, por ejemplo, leyes físicas.
  Condición necesaria pero no suficiente !!!
- La incertidumbre juega un rol importante en la fiabilidad de las relaciones, ya que afecta a **todas** las mediciones.

Objetivo

Introducir técnicas útiles para reconocer la existencia de correlación entre dos cantidades y expresarlo de forma funcional, considerando también la incerteza de las mediciones.

Relaciones entre cantidades físicas

Consideremos cantidades físicas X, Y.

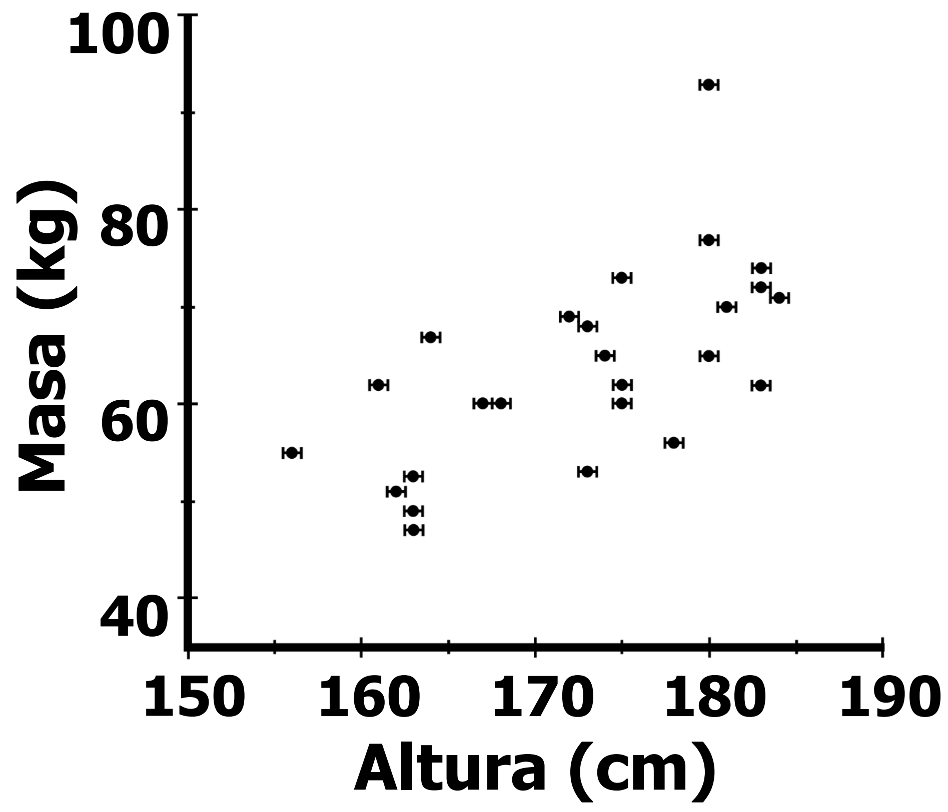
Se mide X N veces, obteniendo valores $x_i, \delta x_i$.

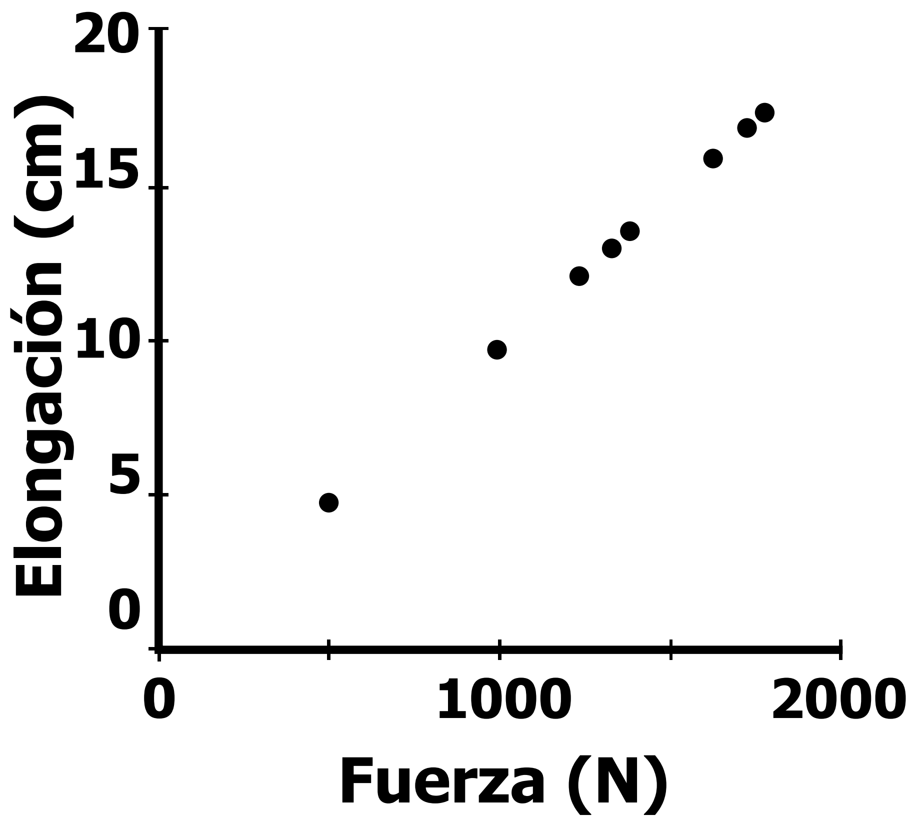
Para X cada se miden $y_i, \delta y_i$.

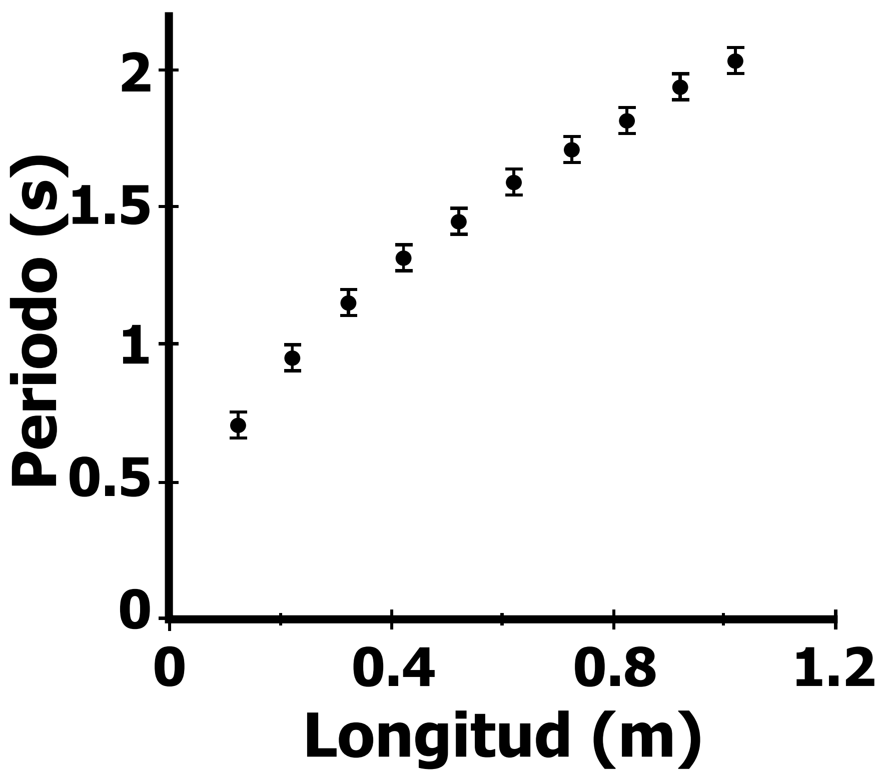
Obtenemos N pares de valores $x_i \pm \delta x_i, \quad y_i \pm \delta y_i, \quad (i = 1, \ldots, N)$

Podemos observar los datos en una tabla o en un gráfico!

Posibles casos





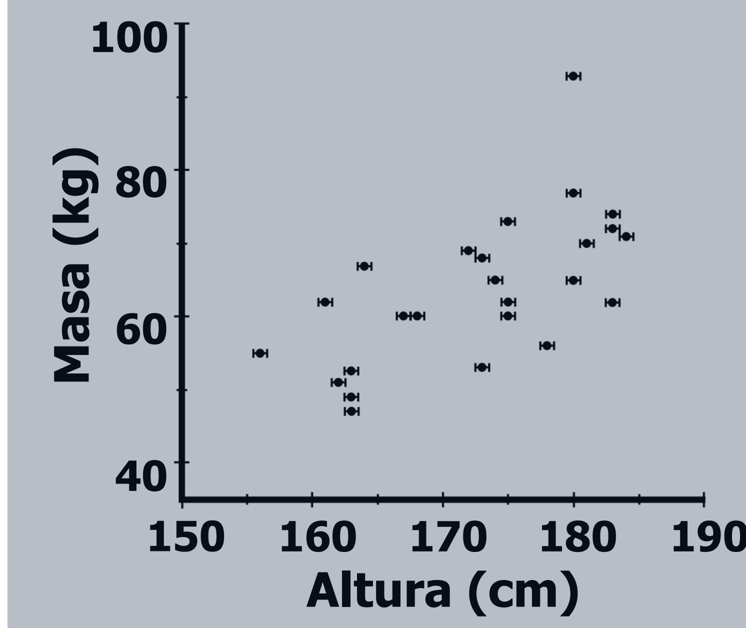


¿Relación lineal?

$$Y = A + BX$$

¿Quiénes son A, B?

¿Relación funcional?



¿Relación lineal?

Varianza y covarianza muestral

N pares (x_j, y_j), $\xi = x, y$

media

desviación estándar

$$m_\xi^* = \frac{1}{N}\sum_j \xi_j, \quad D_\xi^* = \frac{1}{N}\sum_j (\xi_j - m_\xi^*)^2, \quad \sigma_\xi^* = \sqrt{D_\xi^*}$$

varianza (dispersión)

Reminder:
$* =$ de la muestra

La covarianza de la muestra queda $\sigma_{xy}^* = \frac{1}{N}\sum_j (x_j - m_x^*)(y_j - m_y^*)$

Advertencia: Diferencia con el *capítulo 8* respecto a significado de **varianza** y **covarianza**.

Antes: Fluctuaciones aleatorias de dos valores x_j e y_j con sus respectivas medias, conectadas a la incerteza.

Coeficiente de correlación lineal

_r_ para los amigos, o también conocido como coeficiente de Pearson

Para determinar qué tan linealmente correlacionadas están dos magnitudes, usamos *r*:

$$r \equiv \frac{\sigma^*_{xy}}{\sigma^*_x \sigma^*_y} = \frac{\sum_j (x_j - m^*_x)(y_j - m^*_y)}{\sqrt{\sum_j (x_j - m^*_x)^2}\sqrt{\sum_j (y_j - m^*_y)^2}}$$

Y por la desigualdad de Cauchy-Schwartz

$$|\sigma^*_{xy}| \leq \sigma^*_x \sigma^*_y \Rightarrow |r| \leq 1$$

Ejemplo: correlación lineal perfecta

¿Cómo interpretamos el valor de r?

Cantidades $X, Y \longrightarrow Y = A + BX$

Par de valores (x_i, y_i) y medias m_x^*, m_y^*

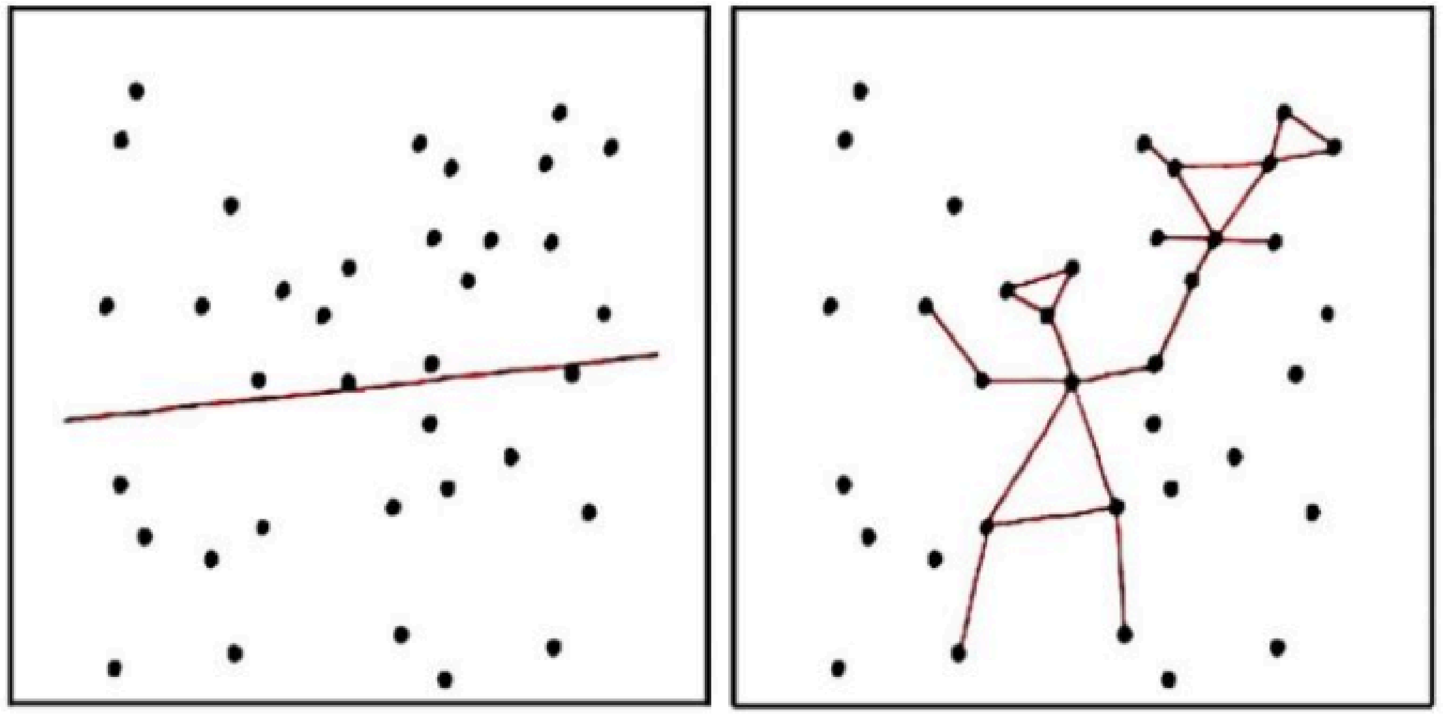
$$\begin{cases} y_i = A + Bx_i \\ m_y^* = A + Bm_x^* \end{cases} \underset{(-)}{\rightarrow} \quad y_i - m_y^* = B(x_i - m_x^*)$$

$$r = \frac{\sum_j (x_j - m_x^*)(y_j - m_y^*)}{\sqrt{\sum_j (x_j - m_x^*)^2}\sqrt{\sum_j (y_j - m_y^*)^2}} = \frac{B}{|B|} = \begin{cases} +1, & B > 0 \\ -1, & B < 0 \end{cases}$$

por eso |*r*|=1 es buena señal cuando hacemos un ajuste !!!

¿$|r|$=1 qué implica?

Y que no …

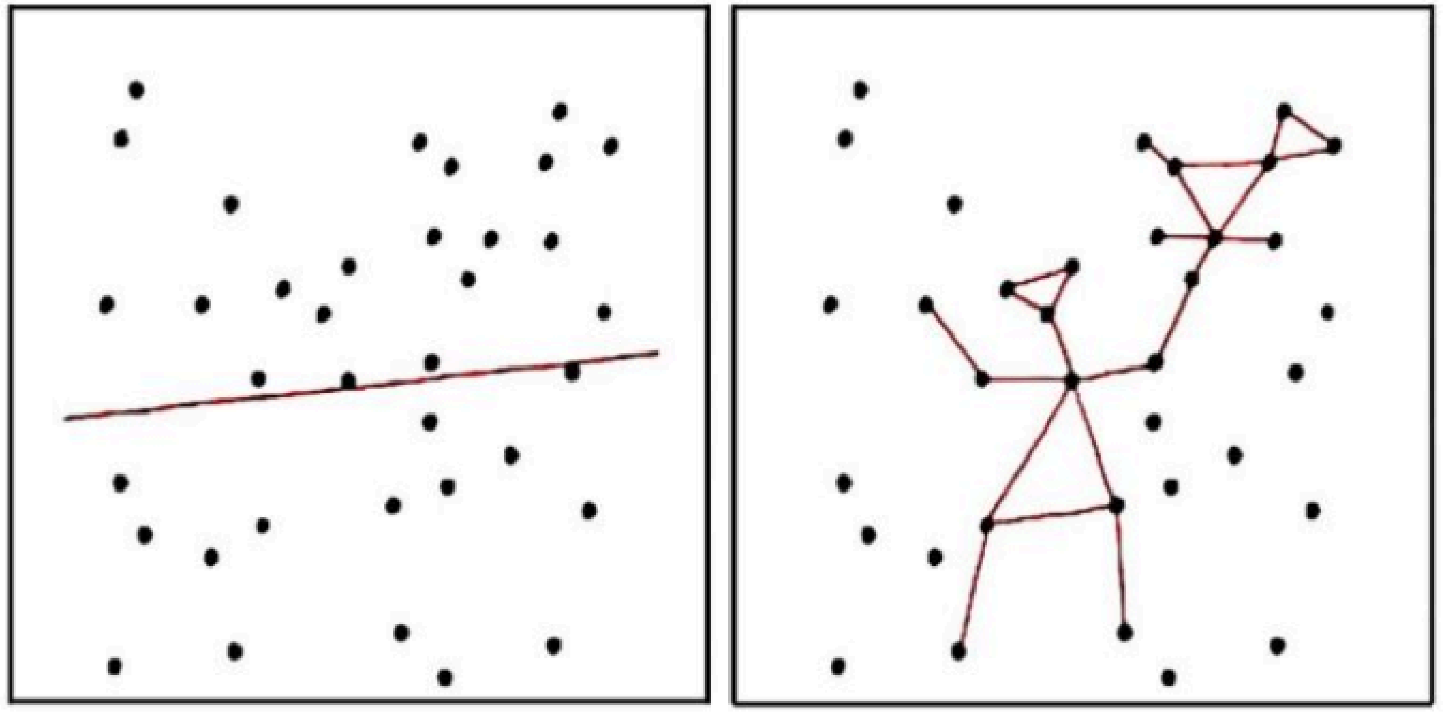


Bestfit Line
Correlation = R = 0.245

The actual best fit
Inspired by Causality

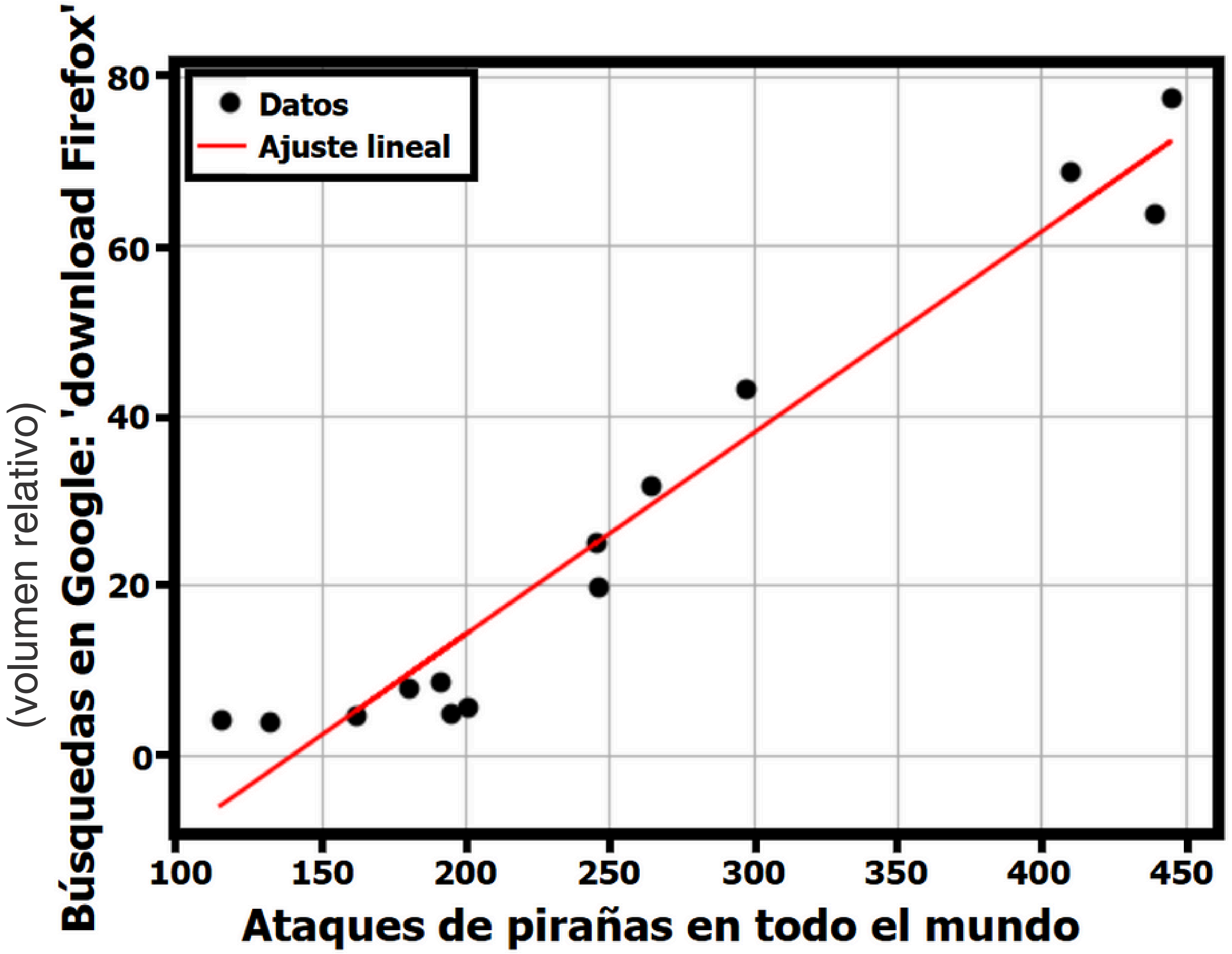
¿|*r*|=1 qué implica?

Y que no ...



Bestfit Line
Correlation = R = 0.245

The actual best fit
Inspired by Causality

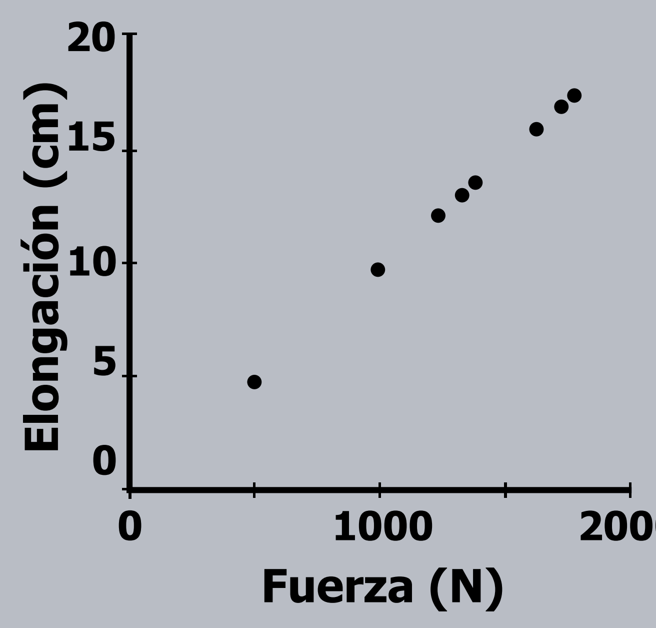


Cada punto representa un año entre 2009 y 2022

r = 0.97

Datos extraídos de *www.tylervigen.com*

Relaciones lineales entre dos cantidades



$Y = A + BX$

¿Quiénes son A, B ?

$$Y = A + BX$$

Tenemos dos objetivos:

- Encontrar A y B → **Regresión lineal**
- Evaluar si, efectivamente, la relación es lineal → **Test de Hipótesis (Capítulo 11)**

Regresión lineal

Basado en cuadrados mínimos

Se quiere minimizar la discrepancia total entre datos y recta.

Hipótesis
δx_i es despreciable

chi cuadrado

$w_i = 1/(\delta y_i)^2$ peso de cada incerteza

$$\chi^2 = \sum_{i=1}^{N} \frac{\overbrace{(y_i - A - Bx_i)^2}^{\text{distancia entre los datos y la recta}}}{(\delta y_i)^2} = \sum_{i=1}^{N} w_i (y_i - A - Bx_i)^2$$

¡Sólo es función de A y B!

$$\chi^2 = \chi^2(A, B) \Rightarrow A, B / \begin{cases} \partial_A \chi^2 = 0 \\ \partial_B \chi^2 = 0 \end{cases}$$

$$\begin{cases} \partial_A \chi^2 = 0 \\ \\ \partial_B \chi^2 = 0 \end{cases} \Rightarrow \begin{cases} (\sum_i w_i) A + (\sum_i w_i x_i) B = \sum_i w_i y_i \\ \\ (\sum_i w_i x_i) A + (\sum_i w_i {x_i}^2) B = \sum_i w_i x_i y_i \end{cases}$$

Sistema 2×2
en A, B

Solución $\rightarrow$

$$\begin{cases} A = \dfrac{(\sum_i w_i {x_i}^2)(\sum_i w_i y_i) - (\sum_i w_i x_i)(\sum_i w_i x_i y_i)}{\Delta_w} \\ \\ B = \dfrac{(\sum_i w_i)(\sum_i w_i x_i y_i) - (\sum_i w_i y_i)(\sum_i w_i x_i)}{\Delta_w} \end{cases}$$

con $\Delta_w = (\sum_i w_i)(\sum_i w_i {x_i}^2) - (\sum_i w_i x_i)^2$

Por suerte las
cuentas las hace
la compu :)

Considerando los errores en *X*

1. Calculamos A', B' considerando sólo error $(\delta y_i)_{\text{experimental}}$

2. Transformamos los errores en X a errores en $Y \longrightarrow (\delta y_i)_{\text{transformada}} = |B'|(\delta x_i)_{\text{exp}}$

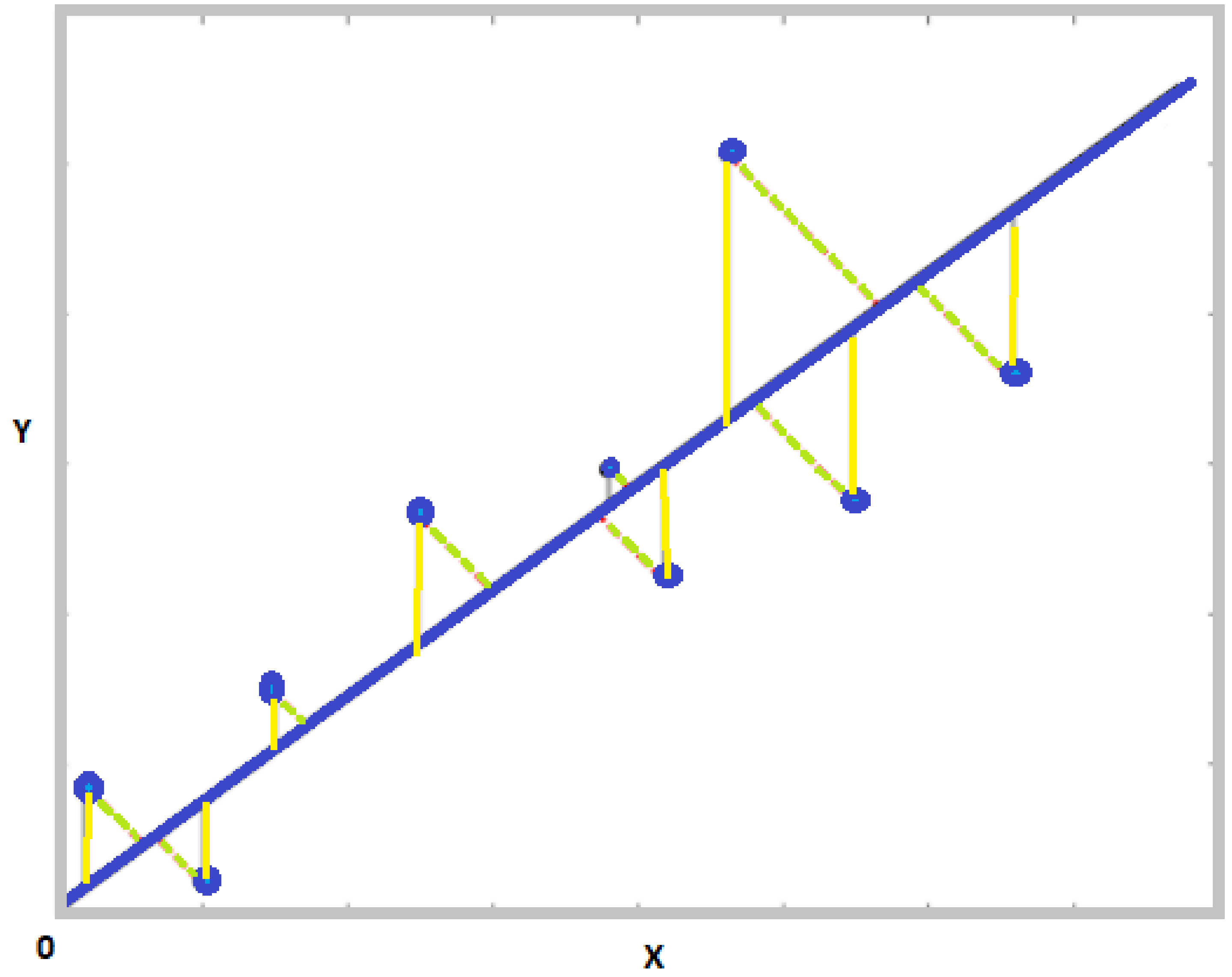
3. Sumamos las contribuciones cuadráticamente $\longrightarrow (\delta y_i)^2_{\text{tot}} = (\delta y_i)^2_{\text{exp}} + (\delta y_i)^2_{\text{tra}}$

4. Calculamos A, B de la misma forma que antes, con $w_i = 1/(\delta y_i)^2_{\text{tot}}$

Pero esto podría no funcionar si los δx_i son grandes

Considerando los errores en *X*: Algoritmo ODR (orthogonal distance regression)

The cool way




```python
import scipy.odr as odr

# Defino función lineal para el ajuste
def lineal(B,x):
    return B[0]*x + B[1] # B[0] = pendiente, B[1] = ordenada

# Configuro modelo en ODR
data = odr.RealData (x, y, sx = error_x, sy = error_y)
modelo = odr.Model(lineal)
odr_instancia = odr.ODR(data, modelo, beta0=[1,1])

# Ejecuto ajuste
resultado = odr_instancia.run()

# Obtengo parámetros ajustados y errores
slope, ordenada = resultado.beta
err_slope, err_ordenada = resultado.sd_beta
```

Errores en *A* y *B*

Recordemos

$$\begin{cases} A &= \dfrac{(\sum_i w_i {x_i}^2)(\sum_i w_i y_i) - (\sum_i w_i x_i)(\sum_i w_i x_i y_i)}{\Delta_w} \\ B &= \dfrac{(\sum_i w_i)(\sum_i w_i x_i y_i) - (\sum_i w_i y_i)(\sum_i w_i x_i)}{\Delta_w} \end{cases} , \quad \Delta_w \neq \Delta_w(y_i)$$

A, B son lineales en y_i!

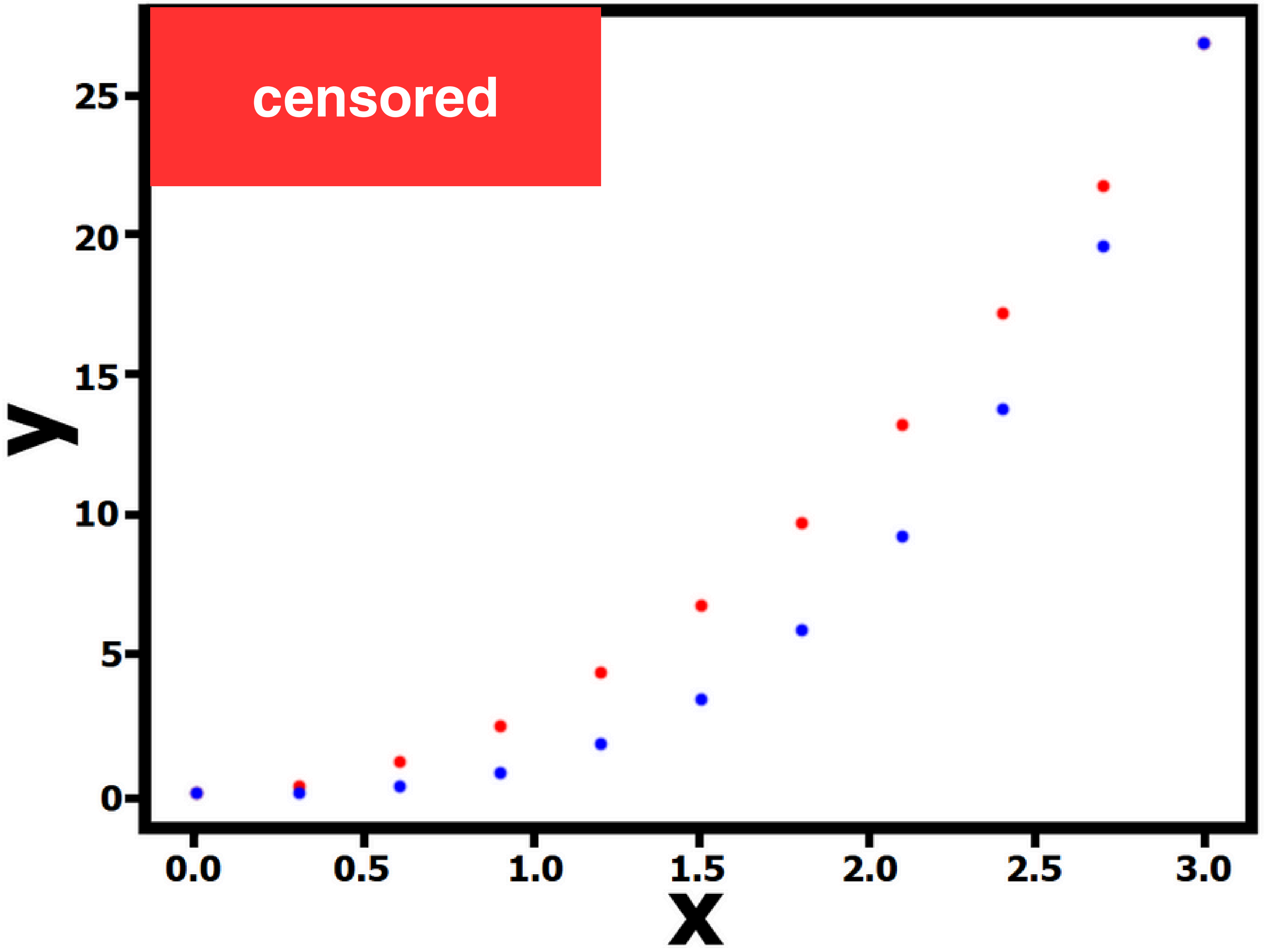
$$\begin{cases} A &= \sum_i \alpha_i y_i \\ B &= \sum_i \beta_i y_i \end{cases}$$

propagando
con $(\delta y_i)^2 = 1/w_i$

$$\begin{cases} (\delta A)^2 &= \sum_i {\alpha_i}^2 (\delta y_i)^2 = \dfrac{\sum_i w_i {x_i}^2}{\Delta_w} \\ (\delta B)^2 &= \sum_i {\beta_i}^2 (\delta y_i)^2 = \dfrac{\sum_i w_i}{\Delta_w} \end{cases}$$

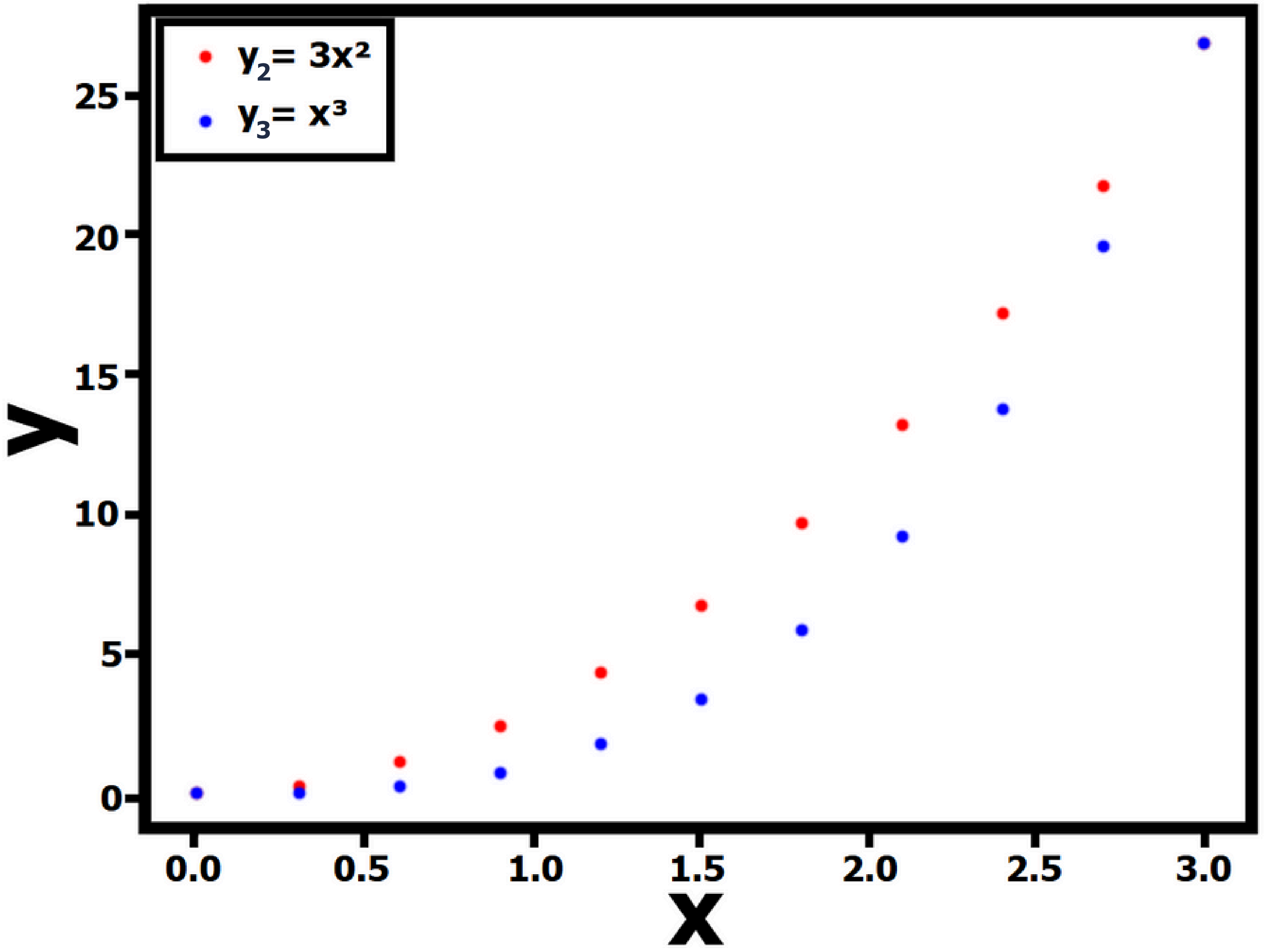
Linealización de relaciones no lineales

¿Cuál es la parábola y cuál la cúbica?



Linealización de relaciones no lineales

¿Cuál es la parábola y cuál la cúbica?



Linealización de relaciones no lineales

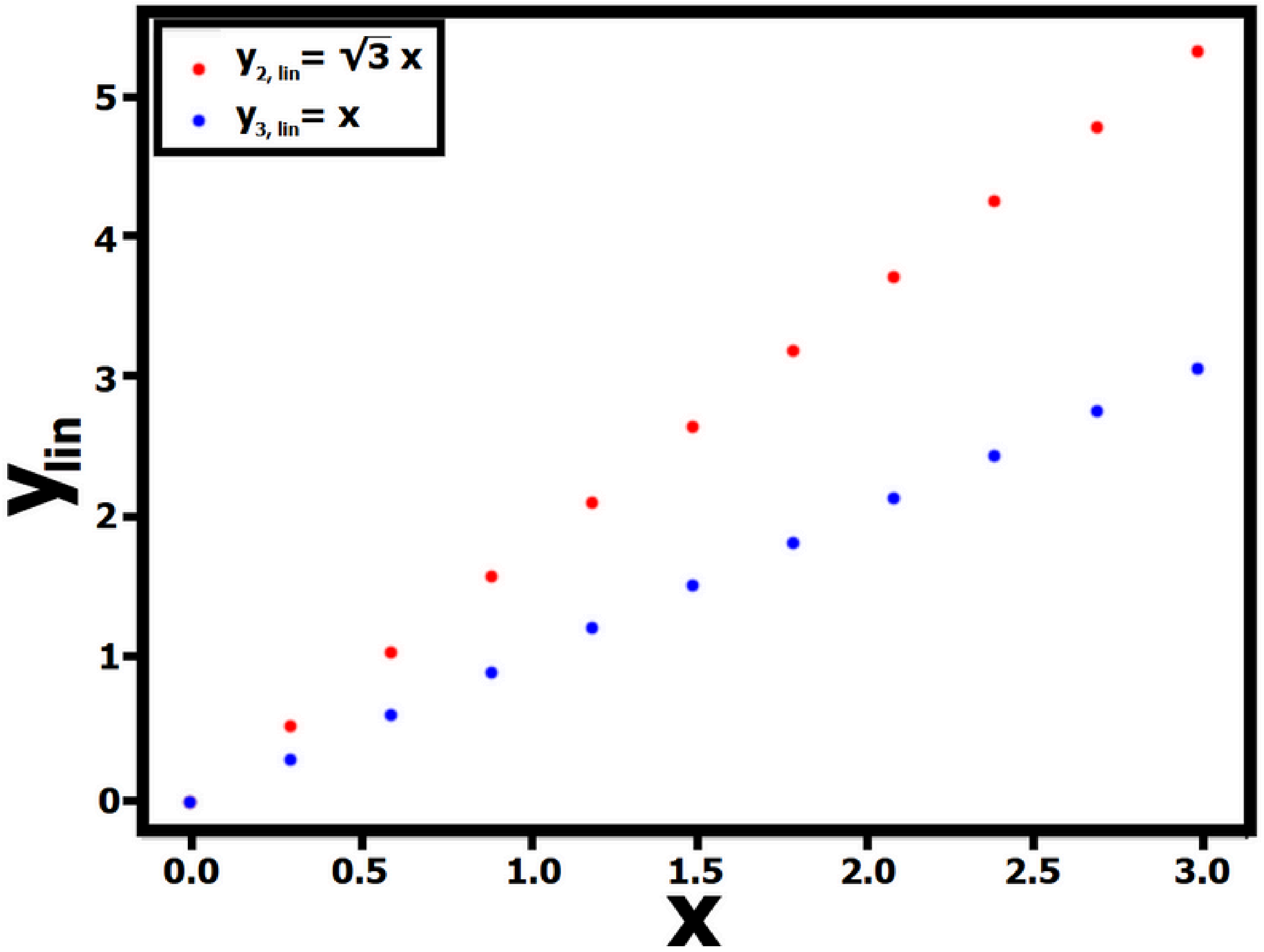
$$Y = \phi(X) \xrightarrow{\quad\quad\quad\quad\quad} Y = A + BX$$

$$Y = \phi^{-1}[\phi(X)]$$

Con el mismo ejemplo

$$\phi_2^{-1} = \sqrt{y_2}$$

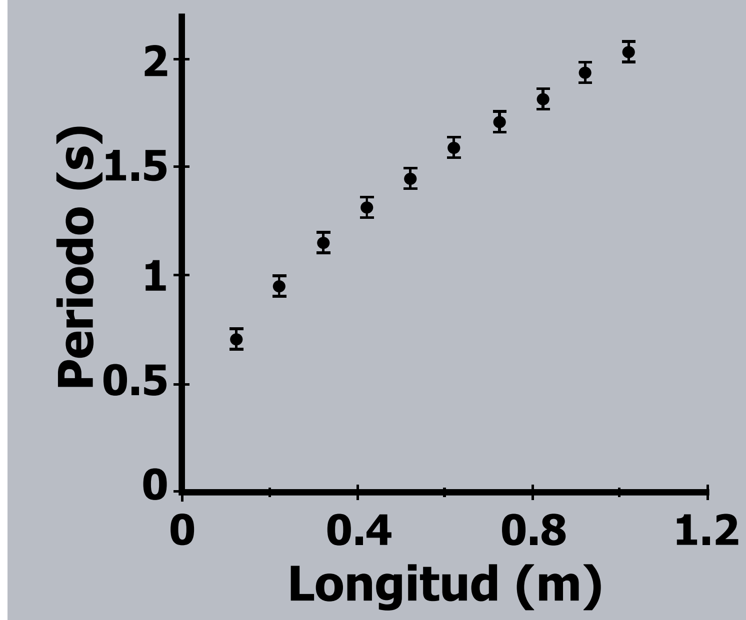
$$\phi_3^{-1} = \sqrt[3]{y_3}$$



Pero la vida, a veces, es no lineal...
(y por eso es difícil)

$$Y = A\exp(-\gamma X)\sin(\omega_0 X + \vartheta)$$

Relaciones entre variables



¿Relación funcional?

Fundamento la regresión lineal con **máxima verosimilitud**

Tengo N pares de valores medidos, de dos magnitudes físicas X, Y:

$$(x_i \pm \delta x_i,\ y_i \pm \delta y_i)$$

SE BUSCA!! $Y = \phi(X, \{\lambda_k\})$ con $\lambda_1, \lambda_2, \ldots, \lambda_k$ parámetros de la función

Ejemplo: $Y = A + BX \longrightarrow Y = \lambda_1 + \lambda_2 X$

$\longrightarrow$ Procedimiento:

1. Definir la forma de la función.
2. Determinar parámetros.

Máxima verosimilitud y cuadrados mínimos

Se deriva del principio de verosimilitud bajo dos hipótesis

- $\delta x_i = 0, \delta y_i \neq 0$
- $\delta y_i = \sigma_i$

desviación estándar $\widetilde{\sigma}_i[m_i^*]$

Dado λ_i, la probabilidad de observar y_i para un dado x_i sigue la distribución:

$$f(y_i) = \frac{1}{\sigma_i\sqrt{2\pi}}\exp\left\{-\frac{[y_i - \phi(x_i, \{\lambda_i\})]^2}{2\sigma_i}\right\}$$

Para N observaciones independientes, la densidad conjunta

$$g(y_1, \ldots, y_N; \{\lambda_k\}) = \frac{1}{(\Pi_i\sigma_i)(2\pi)^{N/2}}\exp\left\{-\frac{1}{2}\overbrace{\sum_{i=1}^{N}\frac{[y_i - \phi(x_i, \{\lambda_k\})]^2}{{\sigma_i}^2}}^{\chi^2}\right\}$$

Minimización para funciones lineales

Cantidad de parámetros λ_k

$$\phi(X, \{\lambda_k\}) = \sum_{k=1}^{p} \lambda_k h_k(X)$$

funciones de *X*

$$\chi^2 = \sum_{i=1}^{N} \frac{[y_i - \phi(x_i, \{\lambda_k\})]^2}{{\sigma_i}^2}$$

Si quiero valores óptimos de λ_k $\frac{\partial \chi^2}{\partial \lambda_k} = \sum_{i=1}^{N} \frac{2}{\sigma_i^2} h_k(x_i) \left[y_i - \sum_{j=i}^{p} \lambda_j h_j(x_i) \right] = 0$

Linealidad de:

- λ_k : Permite analíticamente minimizar el χ^2
- X : Sólo se cumple en regresión lineal simple.

Minimización para funciones no lineales

$Y = ae^{bX} \longrightarrow \ln Y = \ln a + bX \equiv Y' \longrightarrow$ cantidad linealizada

$Y = Ae^{-\gamma X} \sin(\omega_0 X + \vartheta)$

Reminder:
La linealización es en λ

Tomo rangos de valores: $A \in (0, 10)$, $\quad \omega_0 \in (1, 5)$,
$\gamma \in (0, 2)$, $\quad \vartheta \in (0, 2\pi)$.

Elijo un número n de valores posibles para cada parámetro.

Para cada conjunto de datos calculo χ^2.

Obstáculo: Cantidad de combinaciones

Considerando incertezas en *X*

Recordar:

$$\delta x_i = 0, \delta y_i \neq 0$$
$$\delta y_i = \sigma_i$$

Pasos a seguir:

1. Primer ajuste sin considerar incertezas $\longrightarrow \lambda'_k$
2. Propagacion de incertidumbre de X a Y $\longrightarrow (\delta y_i)_{\text{tra}} = \left| \frac{d\phi'(X, \{\lambda'_k\})}{dX} \right|_{x_i} (\delta x_i)_{\text{exp}}$
3. Calculo de incertidumbre total en Y $\longrightarrow (\delta y_i)^2_{\text{tot}} = (\delta y_i)^2_{\text{exp}} + (\delta y_i)^2_{\text{tra}}$
4. Ajuste con incertidumbre corregida $\longrightarrow \sigma_i = (\delta y_i)_{\text{tot}}$

Evaluación a posteriori de la incertidumbre en *Y*

1. Minimizamos la suma de residuos al cuadrado $\psi^2 \equiv \sum_{i=1}^{N} [y_i - \phi(X, \{\lambda_k\}]^2$.

2. Estimamos la incertidumbre media por máxima verosimilitud $\sigma^2 = \dfrac{1}{N} \sum_{i} [y_i - \phi(X, \{\lambda_k\}]^2$.

3. Corrección por grados de libertad $\sigma^2 = \dfrac{1}{N-p} \sum_{i} [y_i - \phi(X, \{\lambda_k\}]^2$.

Gracias :)

¿Consultas?

Escanea el QR para acceder al capítulo, nuestro contacto, y a las diapos!