

Clase 5 - TP2

Comparación de datos con modelos (cuadrados mínimos)

Introducción epistemológica

**Las leyes físicas se expresan
como relaciones funcionales
entre variables cuantitativas**

- Entender / Explicar
- Predecir

2^{da} ley de Newton:

$$\vec{F} = m \cdot \frac{d\vec{v}}{dt}$$

Introducción epistemológica

**Las leyes físicas se expresan
como relaciones funcionales
entre variables cuantitativas**

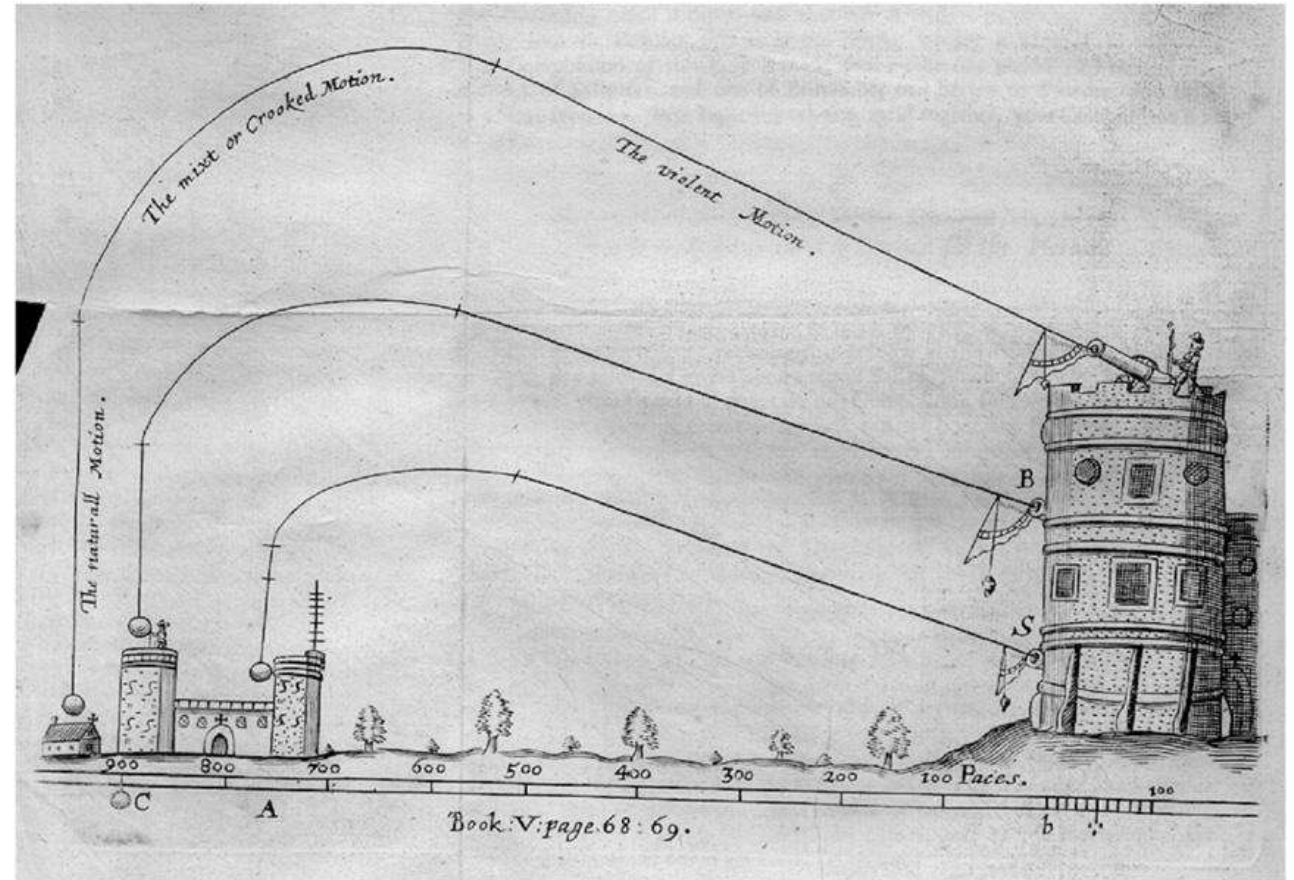
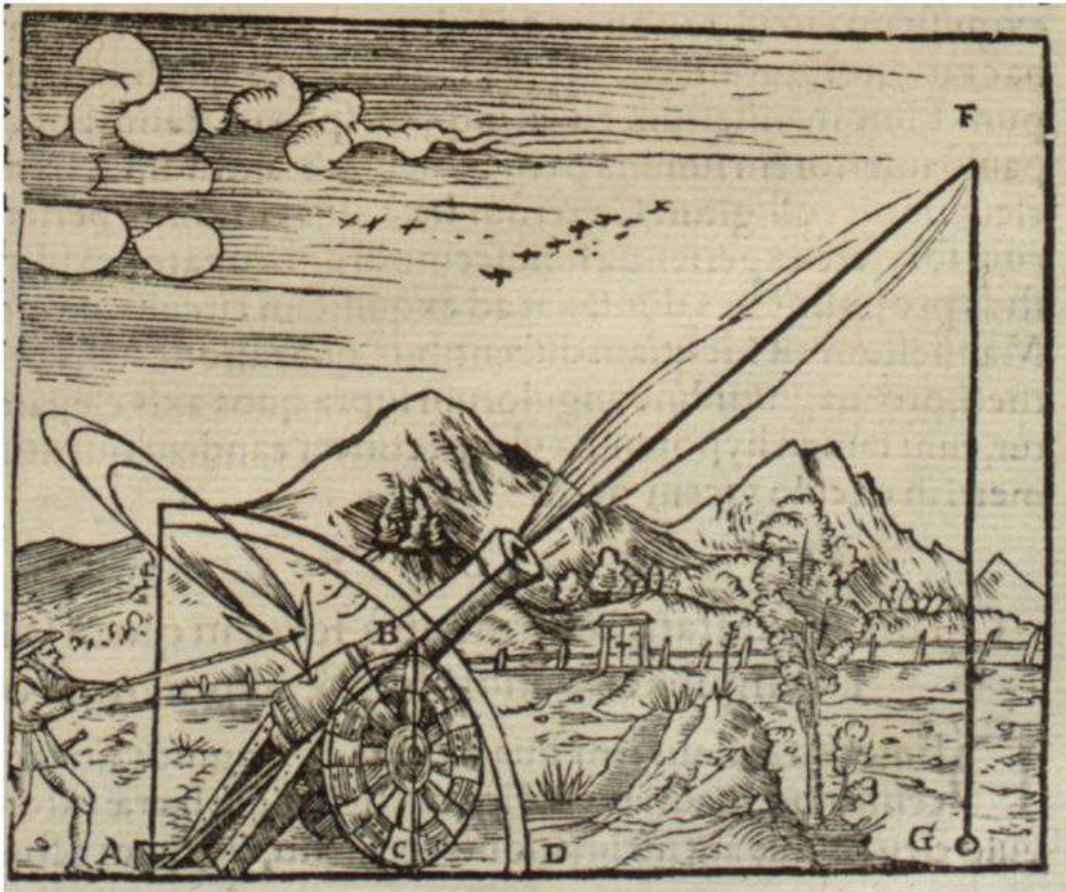
2^{da} ley de Newton:

$$\vec{F} = m \cdot \frac{d\vec{v}}{dt}$$

- Entender / Explicar
- Predecir
- Corroborar → comparar datos con modelos
 - Corroborar validez
 - Inferir parámetros → Medición indirecta

Modelos en física

Representación de tiro oblicuo/balístico en la edad media



Representación de tiro oblicuo/balístico en la edad media

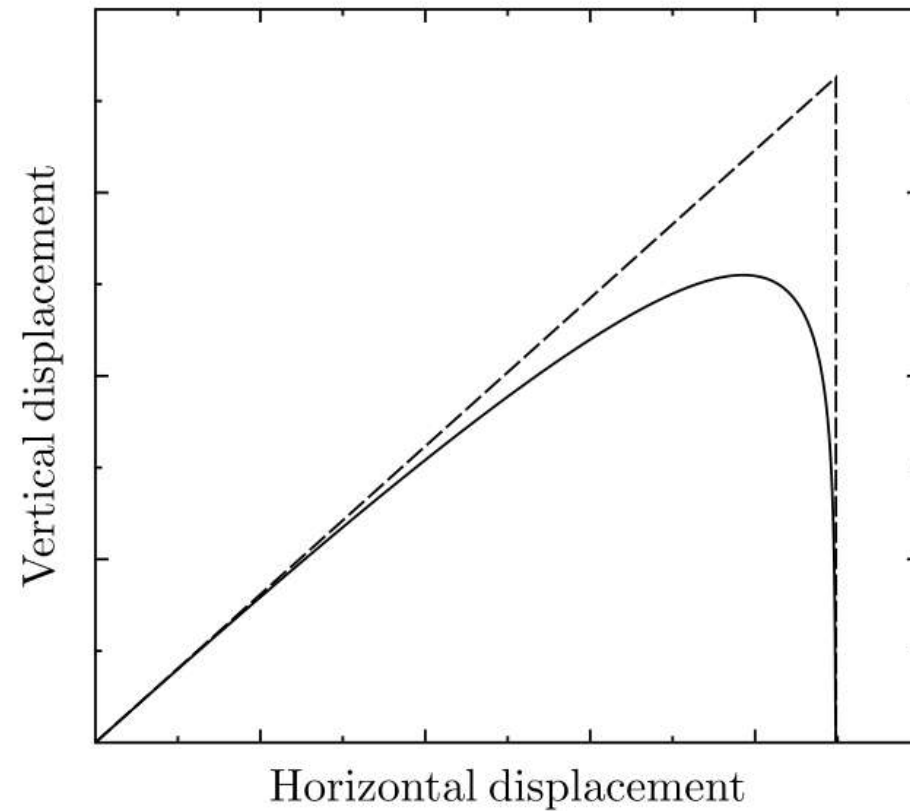
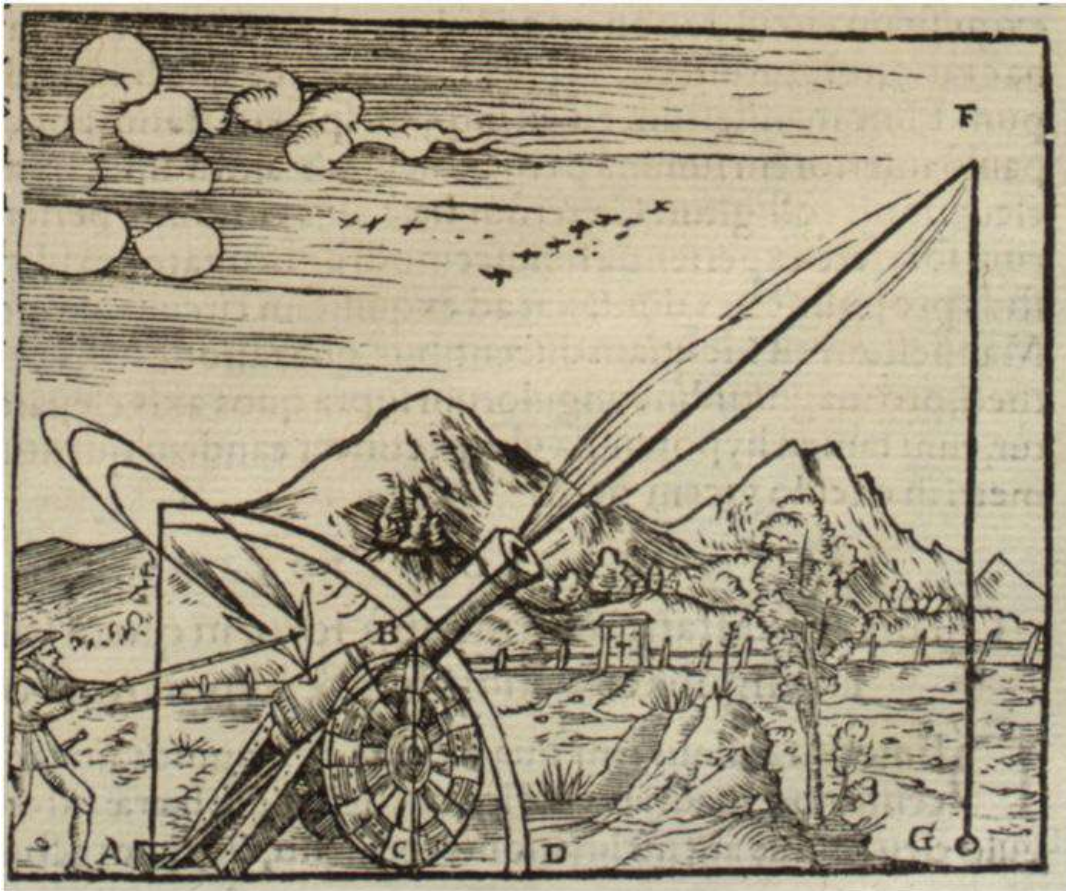
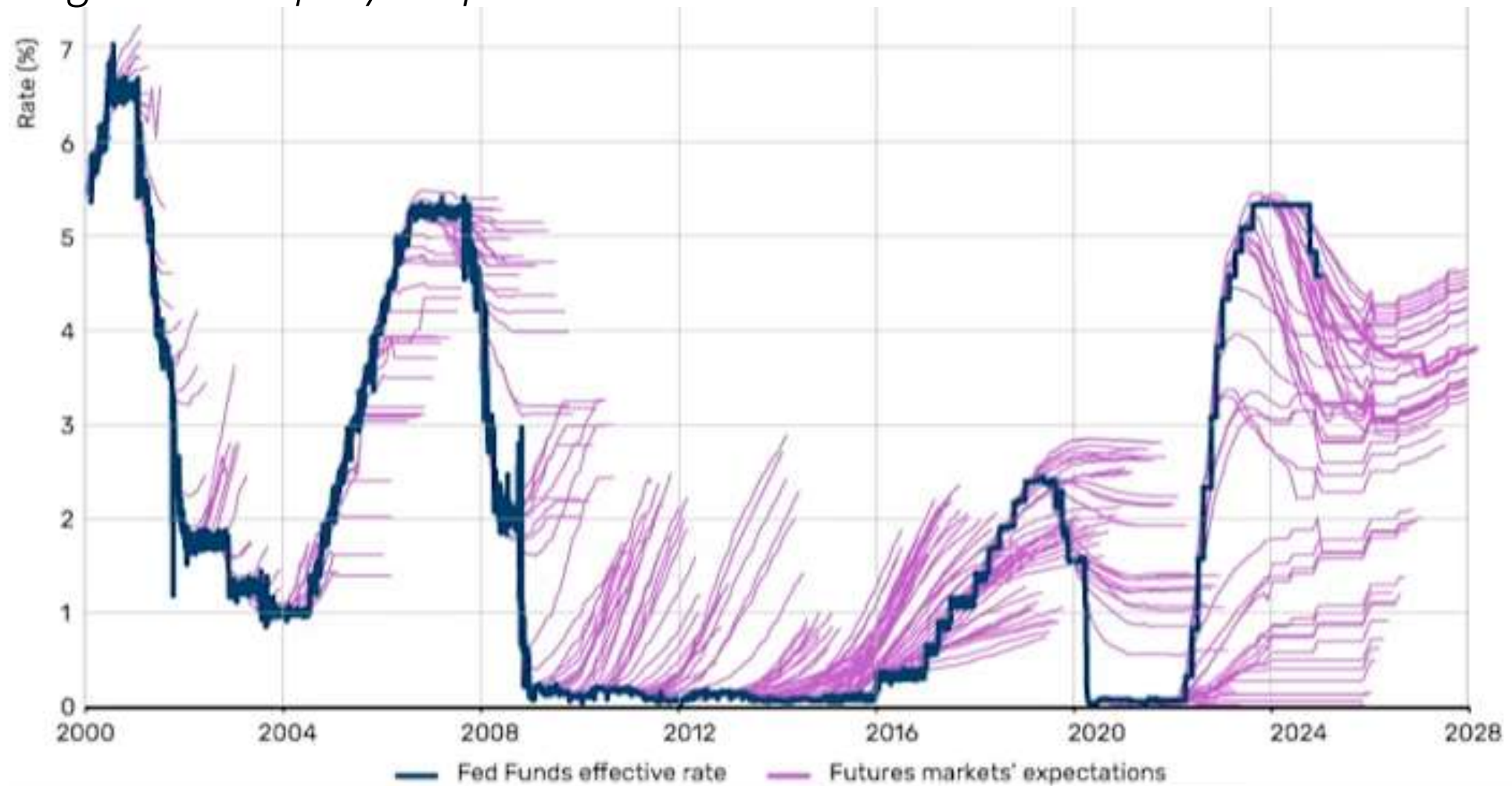


Figure 5. Example of a linearly resisted projectile trajectory in the strong damping limit (solid line) compared to its corresponding approximate triangular trajectory (broken line).

Predicción

"Sin embargo, queríamos principalmente destacar el informe por una versión hermosa y actualizada de uno de los formatos de gráficos favoritos de Alphaville: pronósticos de tasas implícitas a lo largo del tiempo y lo que las tasas realmente han hecho."



Comparación de datos y modelos

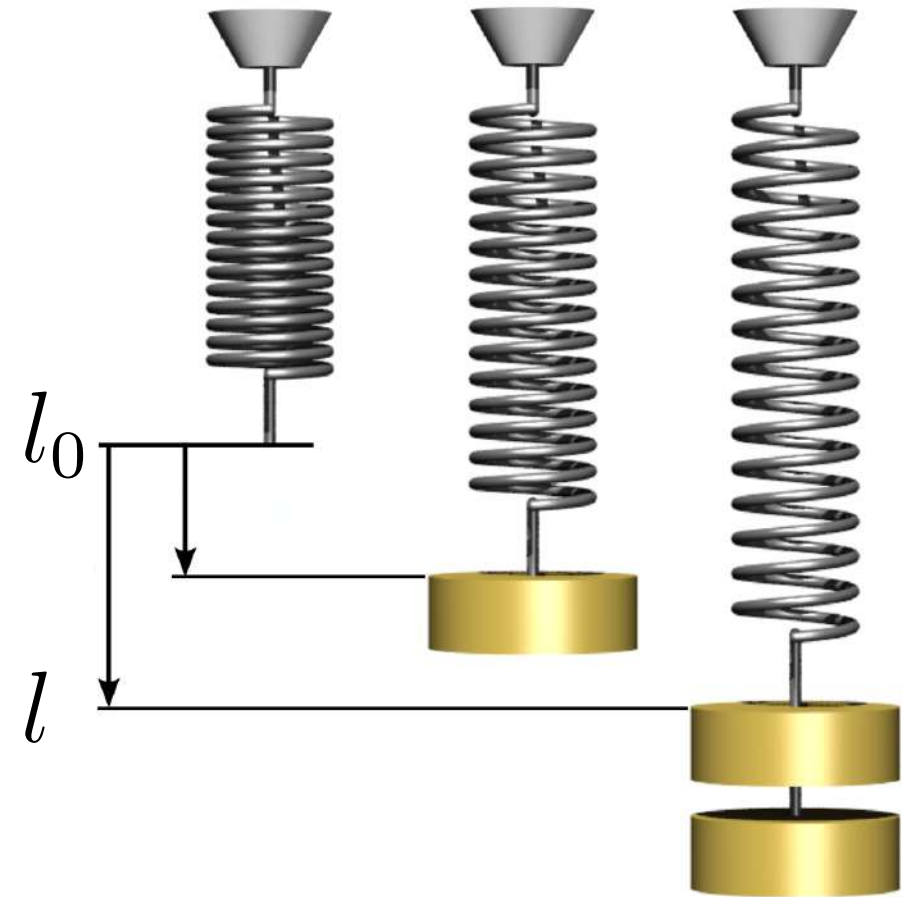
Ejemplo de modelo: Ley de Hook

Relación funcional entre
posición y fuerza: Lineal

$$F = -k \cdot (l - l_0)$$

$$F = l \cdot (-k) + (k \cdot l_0)$$

$$y = x \cdot a + b$$



Cuadrados mínimos

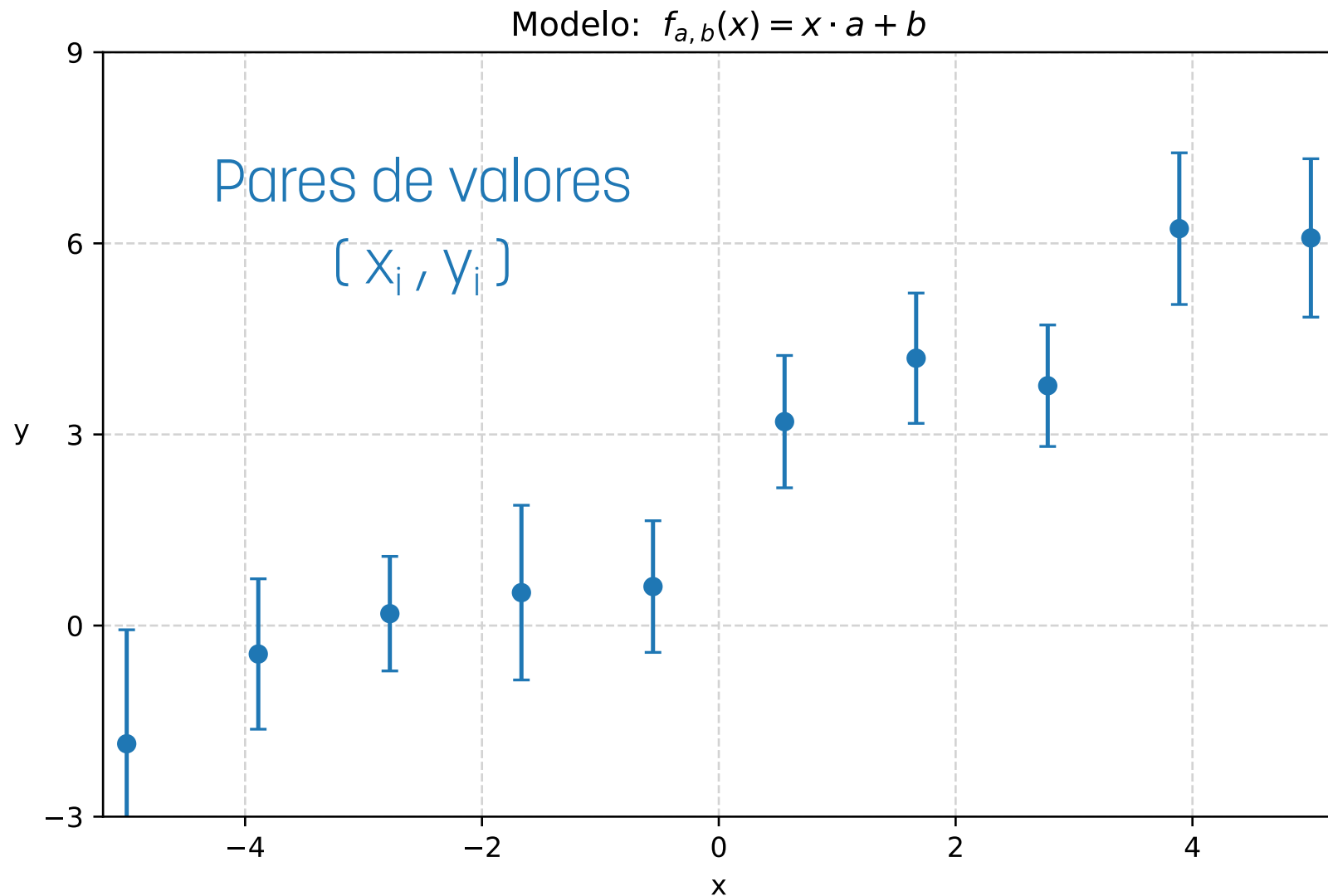
Conjunto de datos medidos (con sus errores)

Modelo Lineal

parámetros

$$f(x) = x \cdot a + b$$

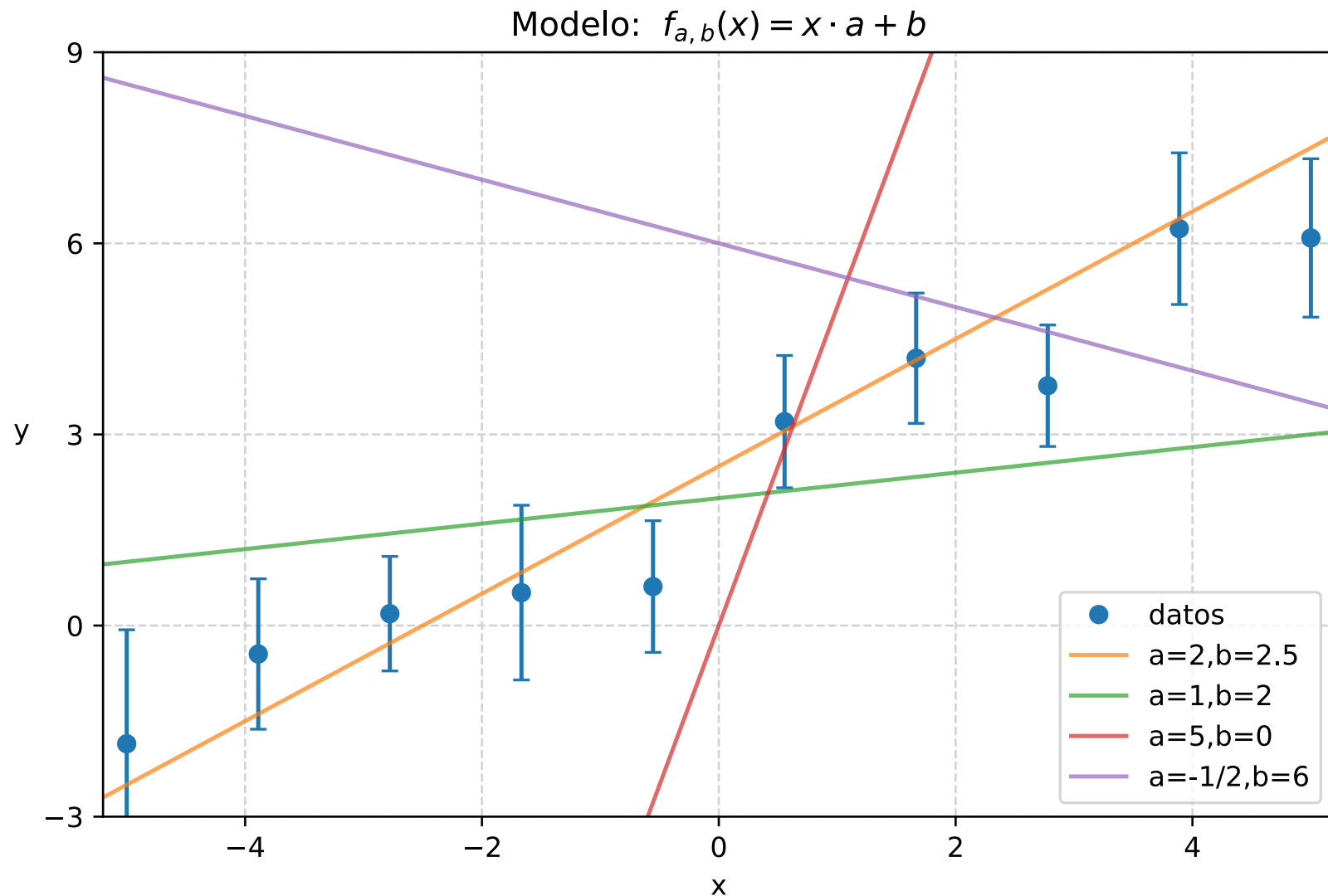
¿Responden
los datos al
modelo?



Cuadrados mínimos

Modelo para diferentes parámetros

¿Con que
parámetros
el modelo
"se parece"
más a los
datos?

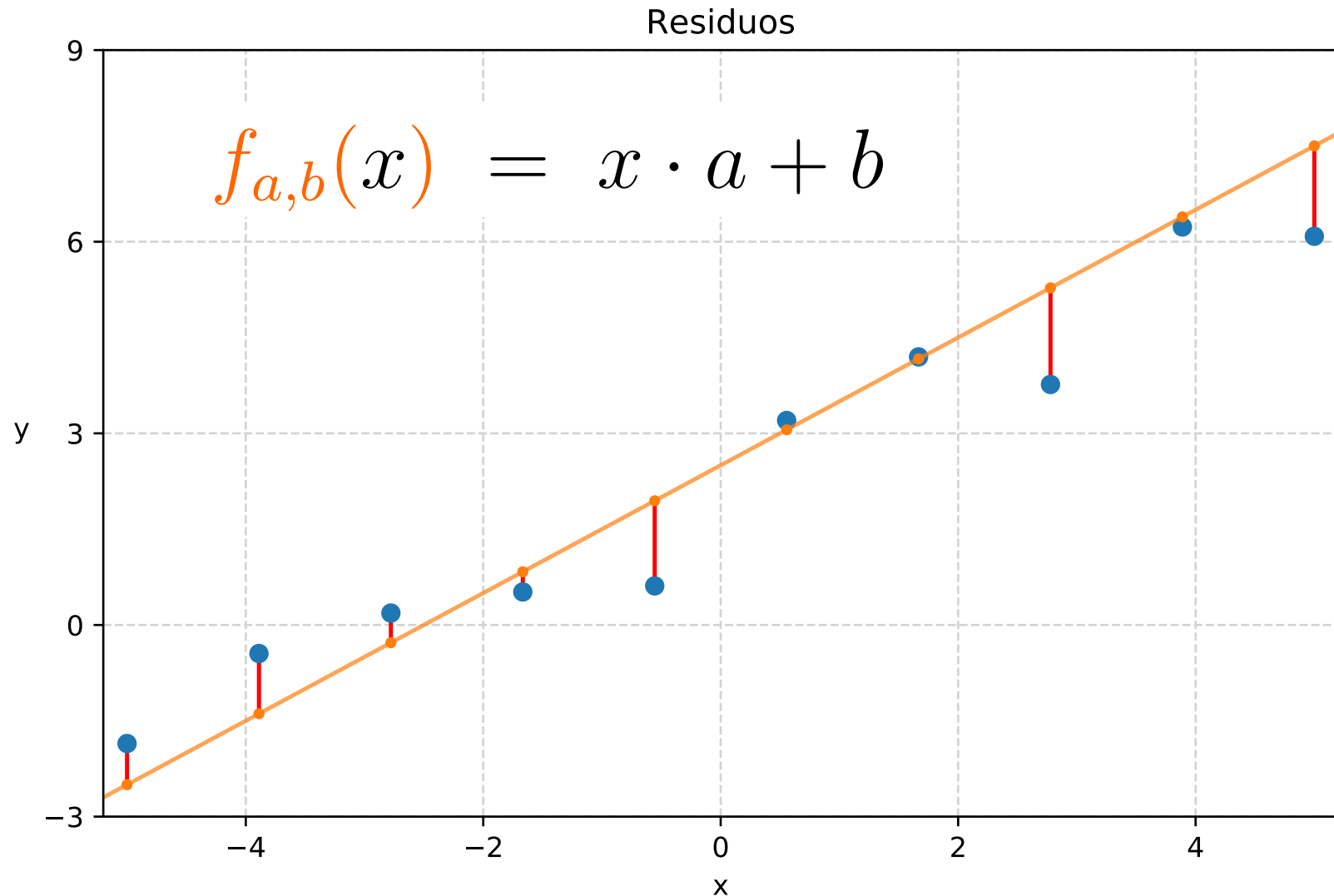


Cuadrados mínimos

$$\text{res}_i = y_i - f_{a,b}(x_i)$$

(x_i, y_i)

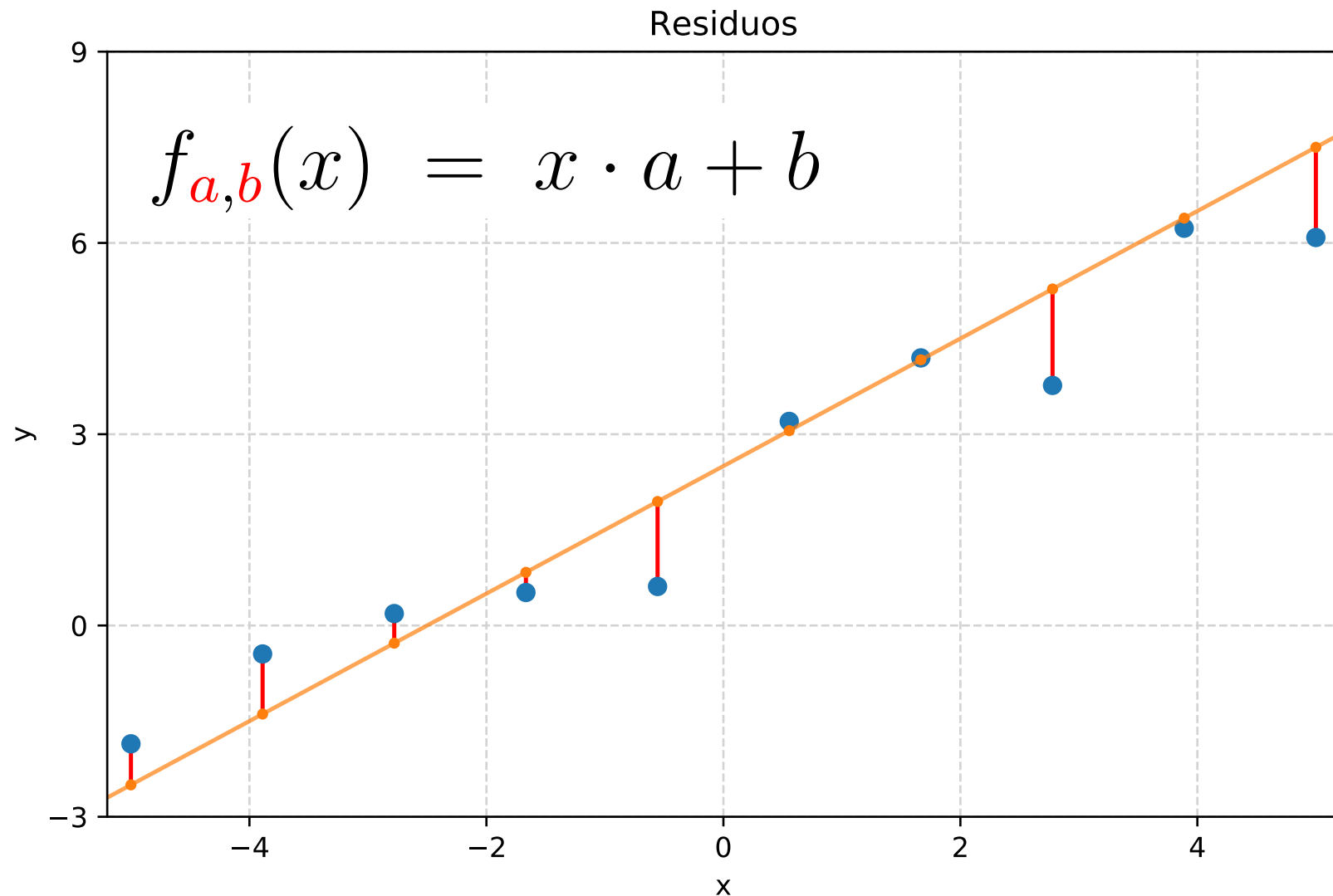
Fabricar una
"medida de
distancia"
entre las
predicciones
del modelo y
los datos
medidos



Cuadrados mínimos

$$S^2(a, b) = \sum_{i=1}^N [y_i - f_{a,b}(x_i)]^2 = \sum_{i=1}^N [y_i - x_i \cdot a - b]^2$$

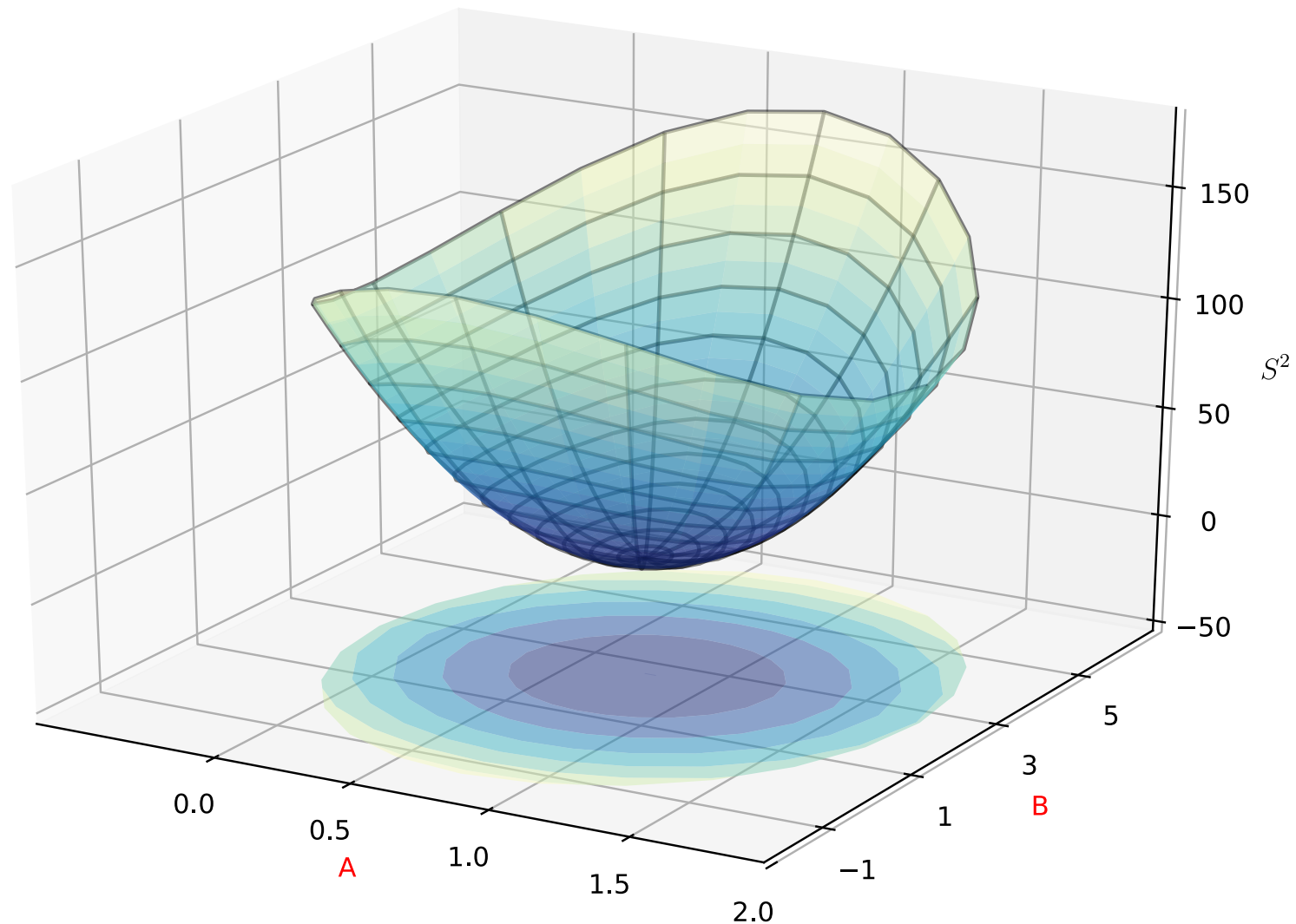
(x_i, y_i)



Cuadrados mínimos

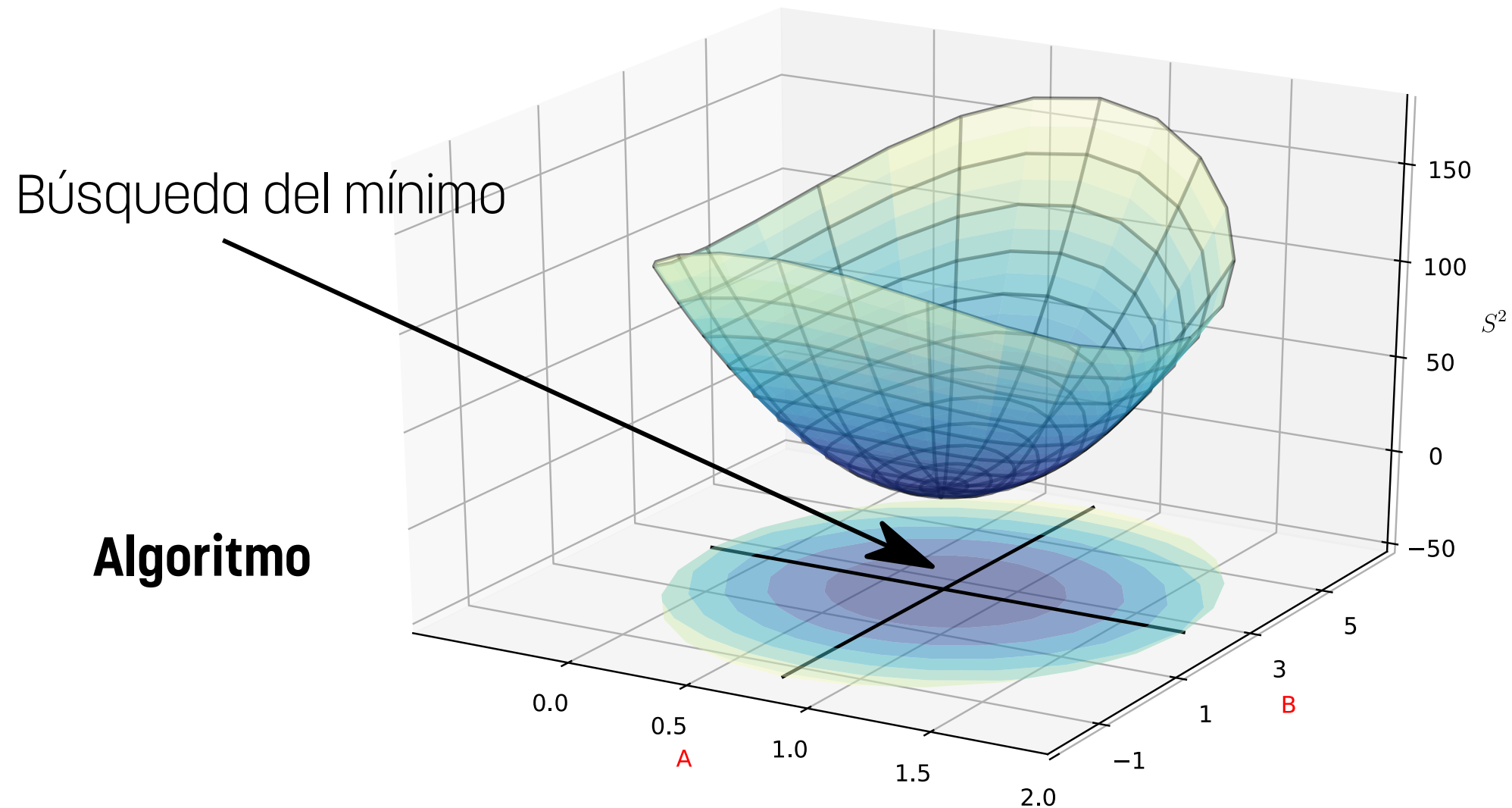
$$S^2(a, b) = \sum_{i=1}^N [y_i - f_{a,b}(x_i)]^2 = \sum_{i=1}^N [y_i - x_i \cdot a - b]^2$$

Superficie que
representa a S^2



Cuadrados mínimos

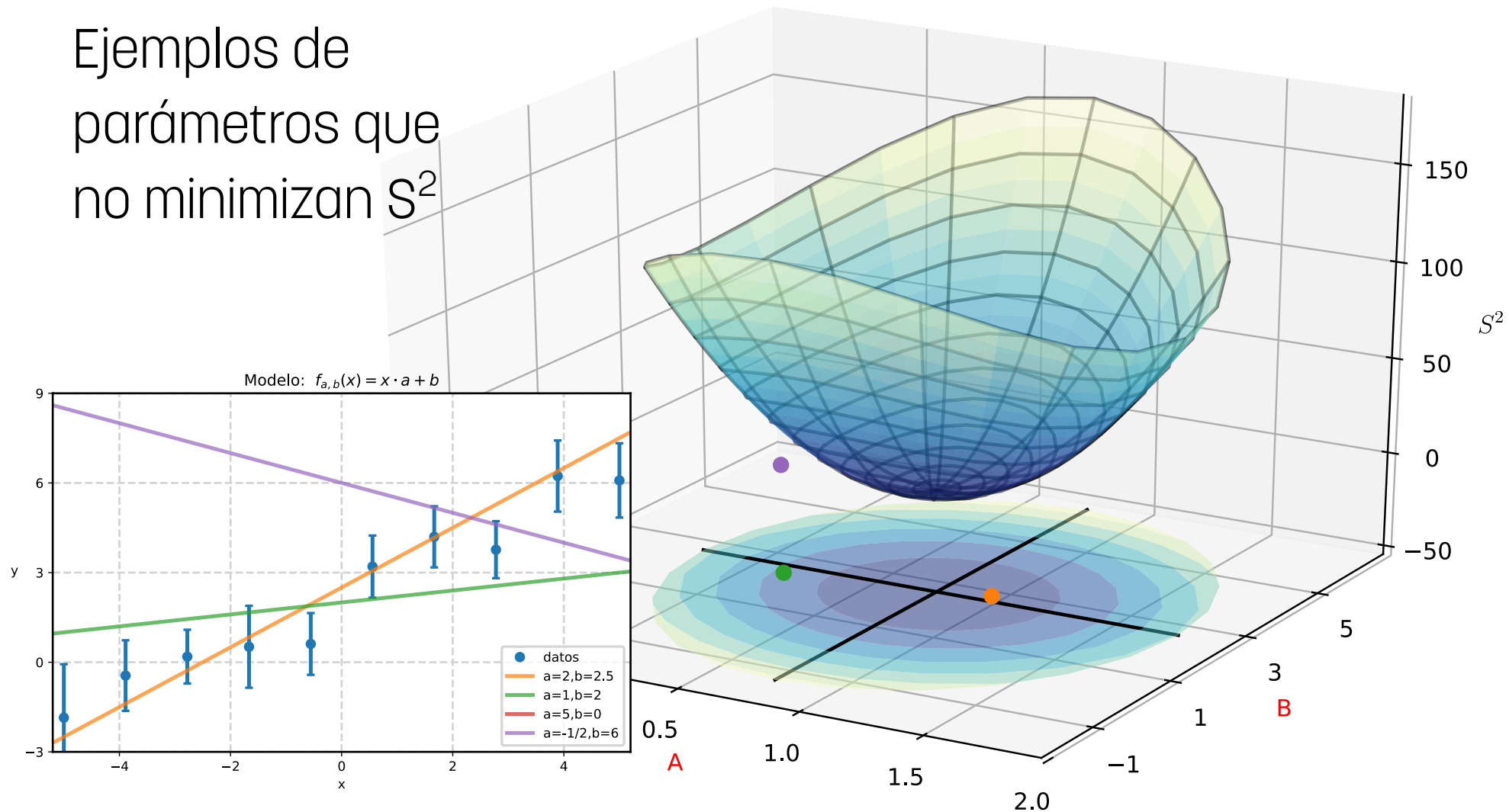
$$S^2(a, b) = \sum_{i=1}^N [y_i - f_{a,b}(x_i)]^2 = \sum_{i=1}^N [y_i - x_i \cdot a - b]^2$$



Cuadrados mínimos

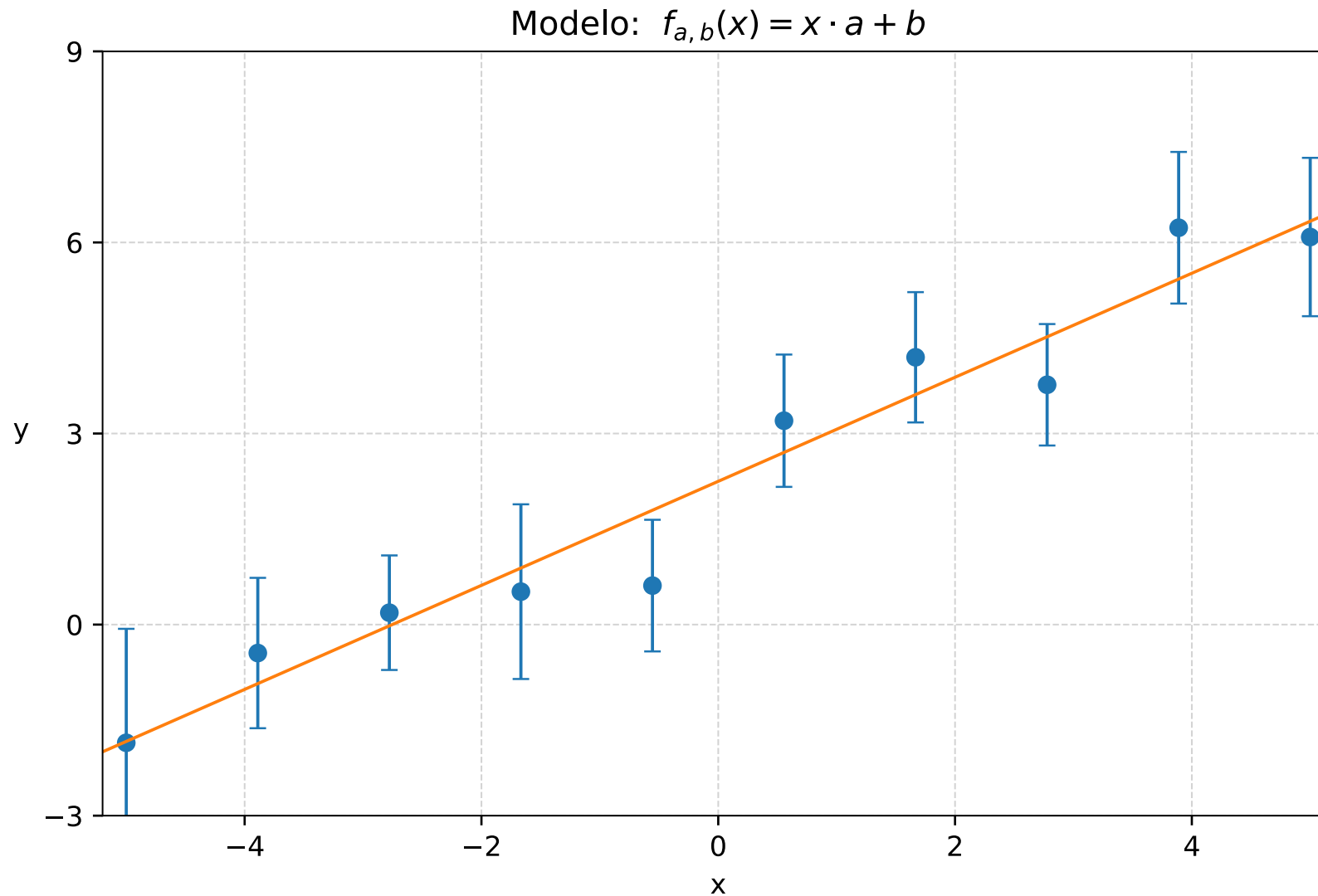
$$S^2(a, b) = \sum_{i=1}^N [y_i - f_{a,b}(x_i)]^2 = \sum_{i=1}^N [y_i - x_i \cdot a - b]^2$$

Ejemplos de
parámetros que
no minimizan S^2



Cuadrados mínimos

Con los parámetros que minimizan S^2
obtenemos el modelo ajustado



Cuadrados mínimos

Algoritmo para modelo lineal (sin incertezas en y)

$$S^2(a, b) = \sum_{i=1}^N [y_i - x_i \cdot a - b]^2$$

$$a = \frac{N \sum (x_i y_i) - (\sum x_i) \cdot (\sum y_i)}{N \sum x_i^2 - (\sum x_i)^2}$$

$$b = \frac{(\sum x_i^2) (\sum y_i) - (\sum x_i) \cdot (\sum x_i y_i)}{N \sum x_i^2 - (\sum x_i)^2}$$

Cuadrados mínimos

Algoritmo para modelo lineal (con incertezas en y)

$$S_w^2(a, b) = \sum_{i=1}^N \left[\frac{y_i - f_{a,b}(x_i)}{\sigma_{y_i}} \right]^2 = \frac{[y_i - f_{a,b}(x_i)]^2}{\sigma_{y_i}^2}$$

$$a = \frac{1}{\Delta} \left[\sum \frac{1}{\sigma_{y_i}^2} \sum \frac{x_i y_i}{\sigma_{y_i}^2} - \sum \frac{x_i}{\sigma_{y_i}^2} \cdot \sum \frac{y_i}{\sigma_{y_i}^2} \right]$$

$$b = \frac{1}{\Delta} \left[\sum \frac{x_i^2}{\sigma_{y_i}^2} \sum \frac{y_i}{\sigma_{y_i}^2} - \sum \frac{x_i}{\sigma_{y_i}^2} \sum \frac{x_i y_i}{\sigma_{y_i}^2} \right]$$

$$\Delta = \sum \frac{1}{\sigma_{y_i}^2} \sum \frac{x_i^2}{\sigma_{y_i}^2} - \left(\sum \frac{x_i}{\sigma_{y_i}^2} \right)^2$$

Estimadores de bondad de ajuste

→ Coeficiente de determinacion, R^2 , r-squared

→ Porcentaje de la varianza que puede explicar el modelo

→ $0 < R^2 < 1$

→ Ej: si $R^2=0.85$, puedo decir que el modelo explica el 85% de la varianza de mis datos

→ Mientras más cercano a 1, mejor es el ajuste

$$R^2 = 1 - \frac{\text{RSS}}{\text{SST}}$$

$$\text{RSS} \equiv S^2 = \sum_i \text{res}_i^2$$

$$\text{SST} \equiv \sum_i (y_i - \langle y \rangle)^2$$

$$R_{adj}^2 = 1 - \frac{\text{RSS}}{\text{SST}} \cdot \frac{N - 1}{N - P - 1}$$

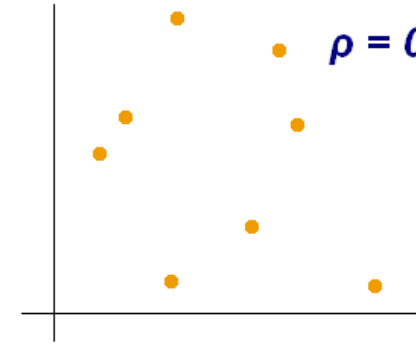
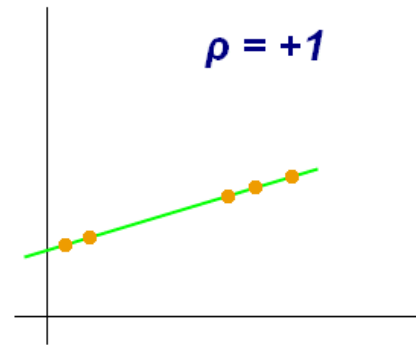
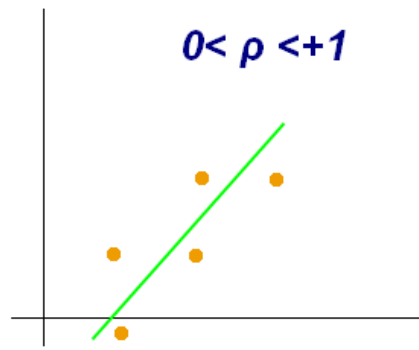
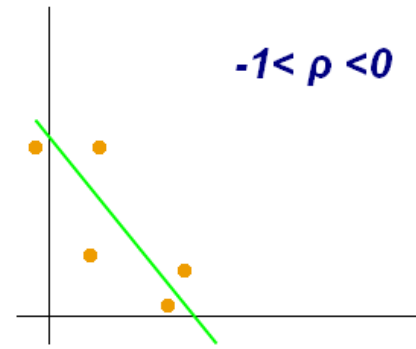
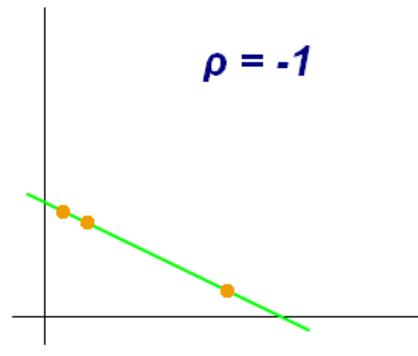
Estimadores de bondad de ajuste

→ Coeficiente de Correlación de Pearson, r , $r\text{-pear}$, ρ

→ "Cuando una variable sube la otra también" o al revés

→ $-1 < \rho < 1$

→ $|\rho| = 1$ significa que los datos están perfectamente alineados



Cuarteto de Anscombe

$$\bar{x} = 9$$

$$\text{std}(x)^2 = 11$$

$$\bar{y} = 7.50$$

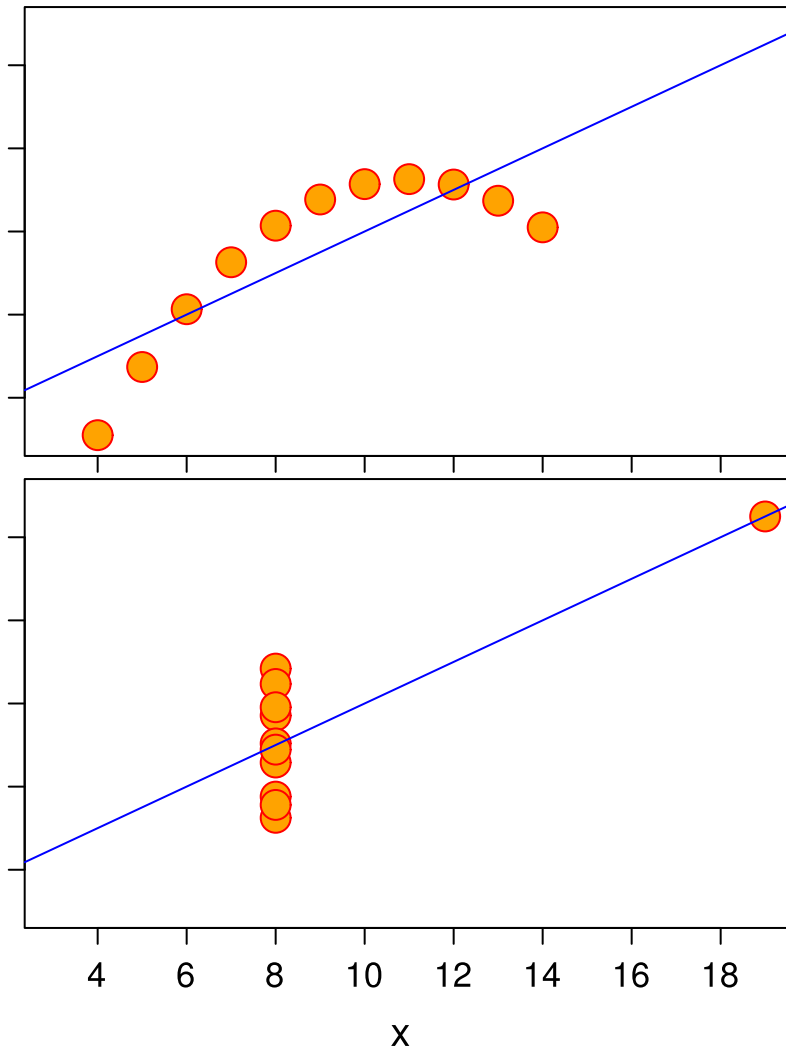
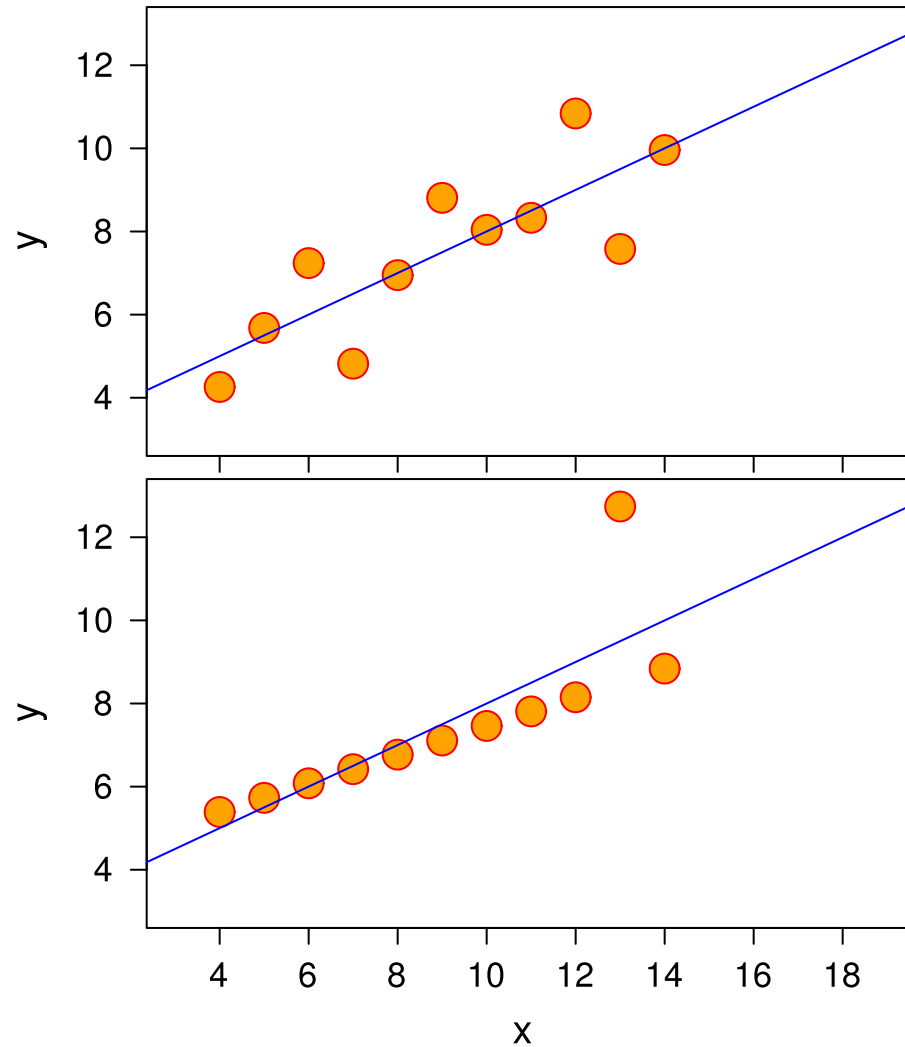
$$\text{std}(y)^2 = 4.125$$

$$\text{cor}(x, y) = 0.816$$

$$R^2 = 0.67$$

linear fit

$$y = 3.00 + 5.000 \cdot x$$

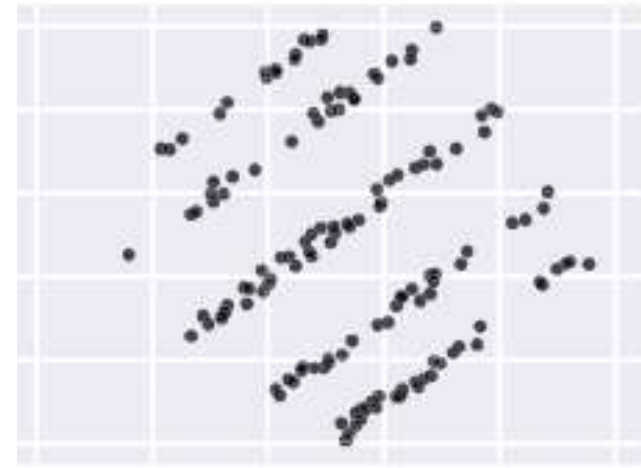
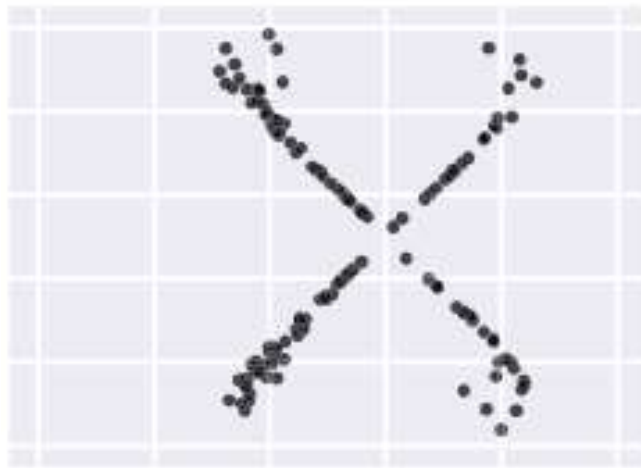
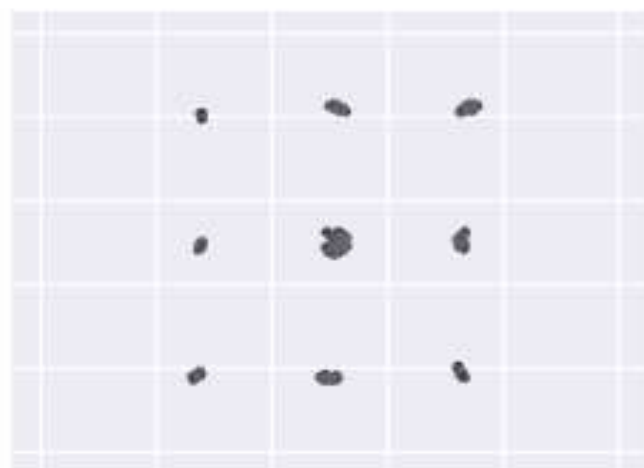
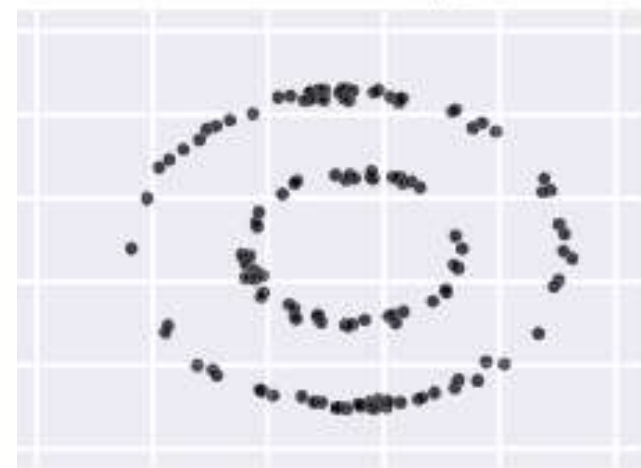
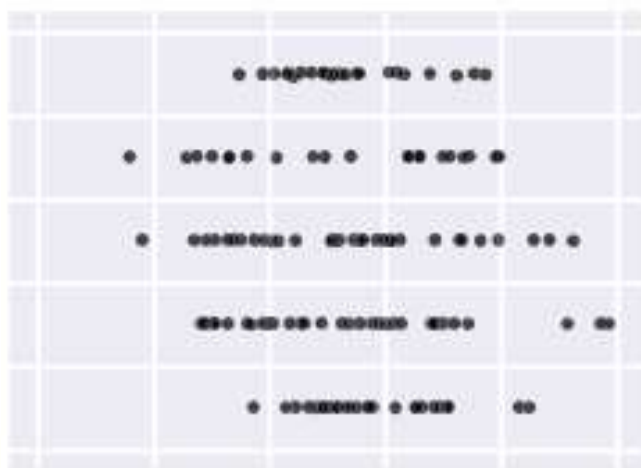
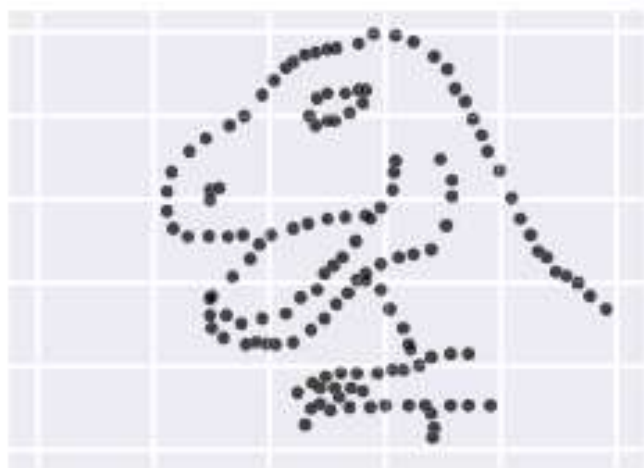


Datasaurus dozen

https://en.wikipedia.org/wiki/Datasaurus_dozen



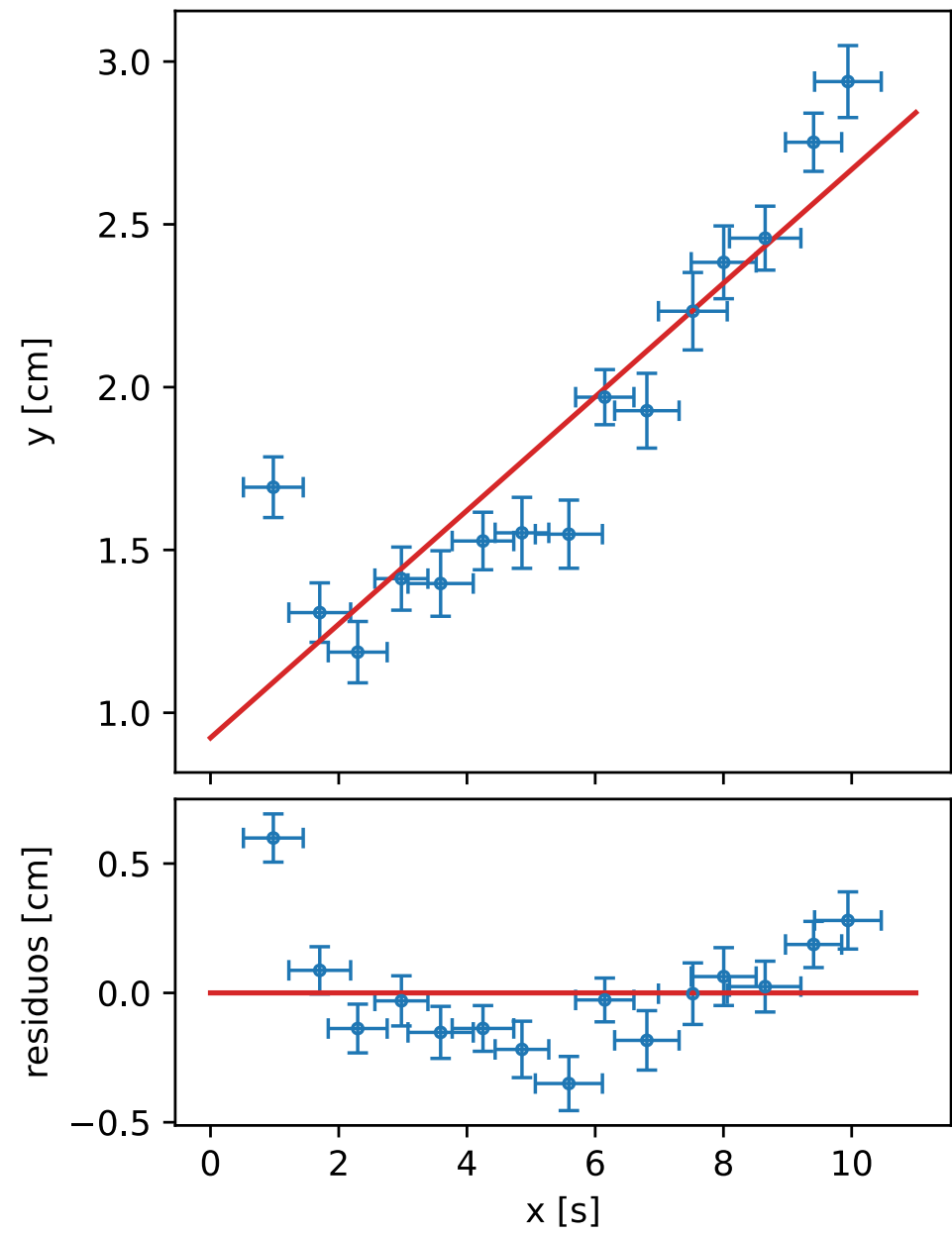
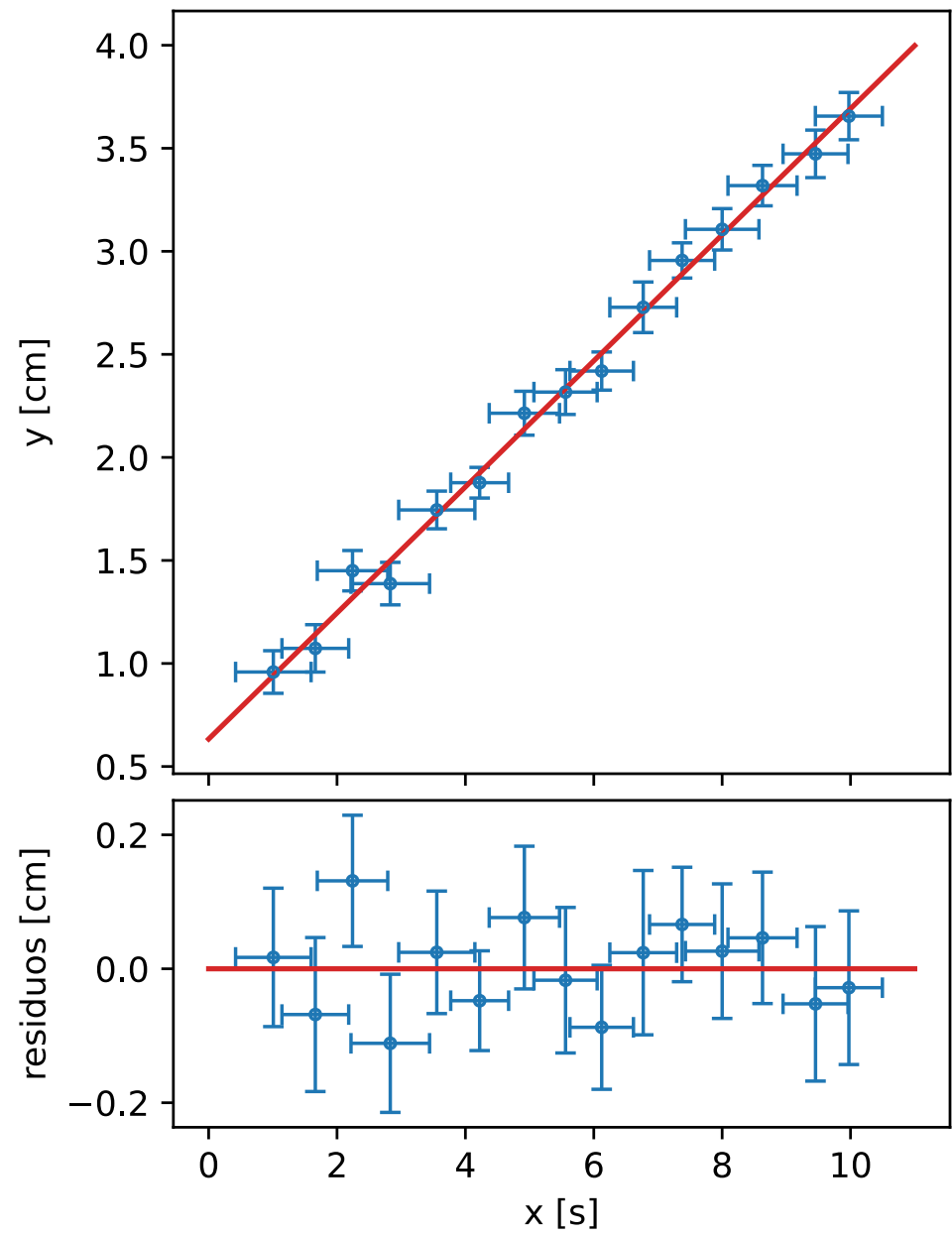
YouTube



**Correlación NO
implica
causalidad**



Residuos



¿...y si el modelo no es lineal?

→ **Usar cantidades que sí tengan relación lineal**

Ej: Caída libre por aceleración de la gravedad

$$y(t) = y_0 + t \cdot v_i + t^2 \cdot \frac{1}{2}g$$

$$y(t) = t^2 \cdot \left(\frac{1}{2}g\right)$$

A diagram illustrating the linear relationship $Y = X \cdot a$. The equation is enclosed in a black rectangular box. From the right side of the box, two arrows originate. The upper arrow points to the definition $a \equiv \frac{1}{2}g$, and the lower arrow points to the definition $X \equiv t^2$.

$$Y = X \cdot a$$
$$a \equiv \frac{1}{2}g$$
$$X \equiv t^2$$

¿...y si el modelo no es lineal?

→ Usar cantidades transformadas (ej: aplicar logaritmo)

$$y(t) = t^2 \cdot \left(\frac{1}{2}g\right)$$

$$\log(y) = \log\left(t^2 \cdot \frac{1}{2}g\right)$$

$$\log(y) = 2 \cdot \log(t) + \log\left(\frac{1}{2}g\right)$$

$$Y = a \cdot X + b$$

$$a \equiv 2$$

$$b \equiv \log\left(\frac{1}{2}g\right)$$

$$X \equiv \log(t)$$

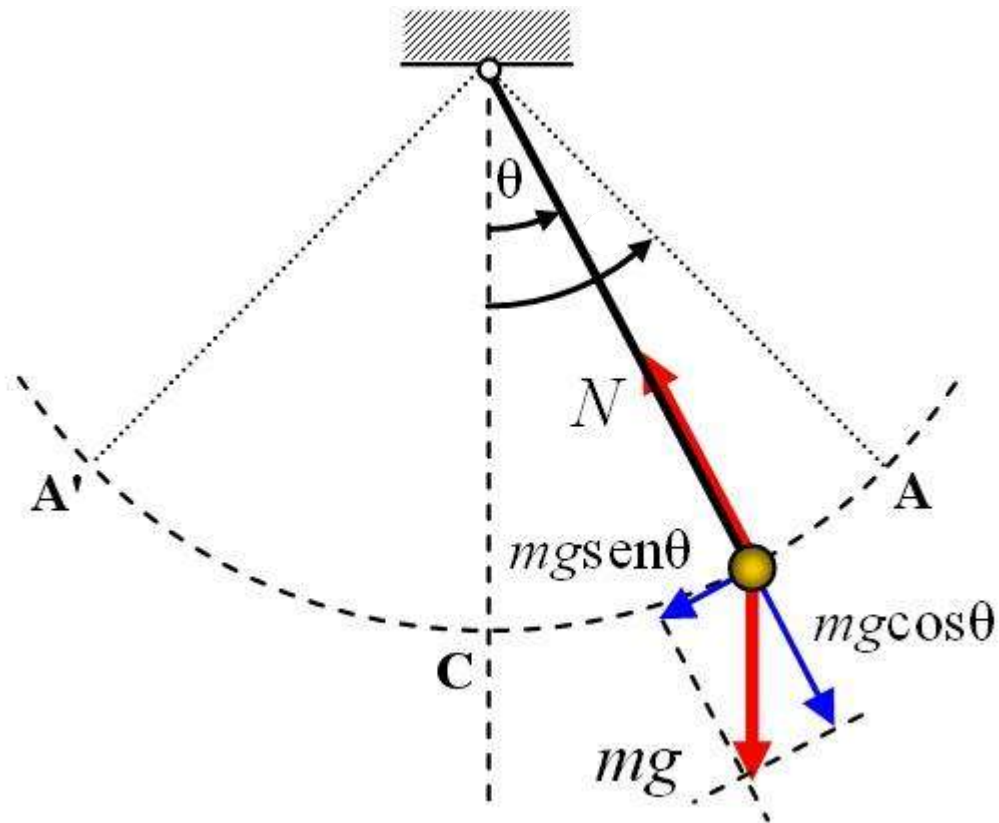
$$Y \equiv \log(y)$$

Actividad

- Medición indirecta de g usando un péndulo
- Evaluar validez del modelo de péndulo simple
- Ajuste de datos con regresión lineal por cuadrados mínimos
- Informe completo

Actividad: validar modelo péndulo

Dinámica de un péndulo



Dinámica:

$$F_t = -m \cdot g \cdot \sin(\theta) = m \cdot L \cdot \ddot{\theta}$$

Ecuación diferencial:

$$L \cdot \ddot{\theta} + g \cdot \sin(\theta) = 0$$

Aproximación: $\sin(\theta) \sim \theta$ $\theta \ll 1$

$$L \cdot \ddot{\theta} + g \cdot \theta = 0$$

Solución:

$$\theta(t) = A \cdot \sin(\omega t + \phi)$$

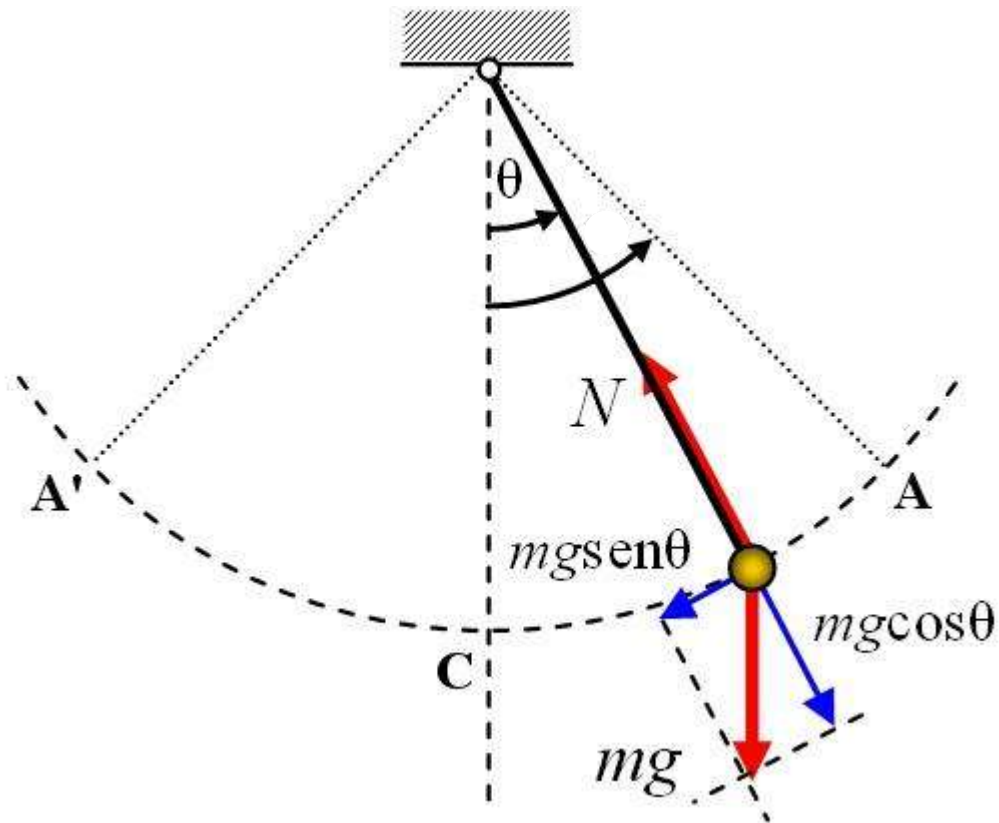
$$\omega = \frac{2\pi}{T} = \sqrt{\frac{g}{L}}$$

Modelo a validar:

$$T = 2\pi \sqrt{\frac{L}{g}}$$

Actividad: validar modelo péndulo

Dinámica de un péndulo



Dinámica:

$$F_t = -m \cdot g \cdot \sin(\theta) = m \cdot L \cdot \ddot{\theta}$$

Ecuación diferencial:

$$L \cdot \ddot{\theta} + g \cdot \sin(\theta) = 0$$

Aproximación: $\sin(\theta) \sim \theta$ $\theta \ll 1$

$$L \cdot \ddot{\theta} + g \cdot \theta = 0$$

Solución:

$$\theta(t) = A \cdot \sin(\omega t + \phi)$$

¿Hipótesis del modelo?

$$\omega = \frac{2\pi}{T} = \sqrt{\frac{g}{L}}$$

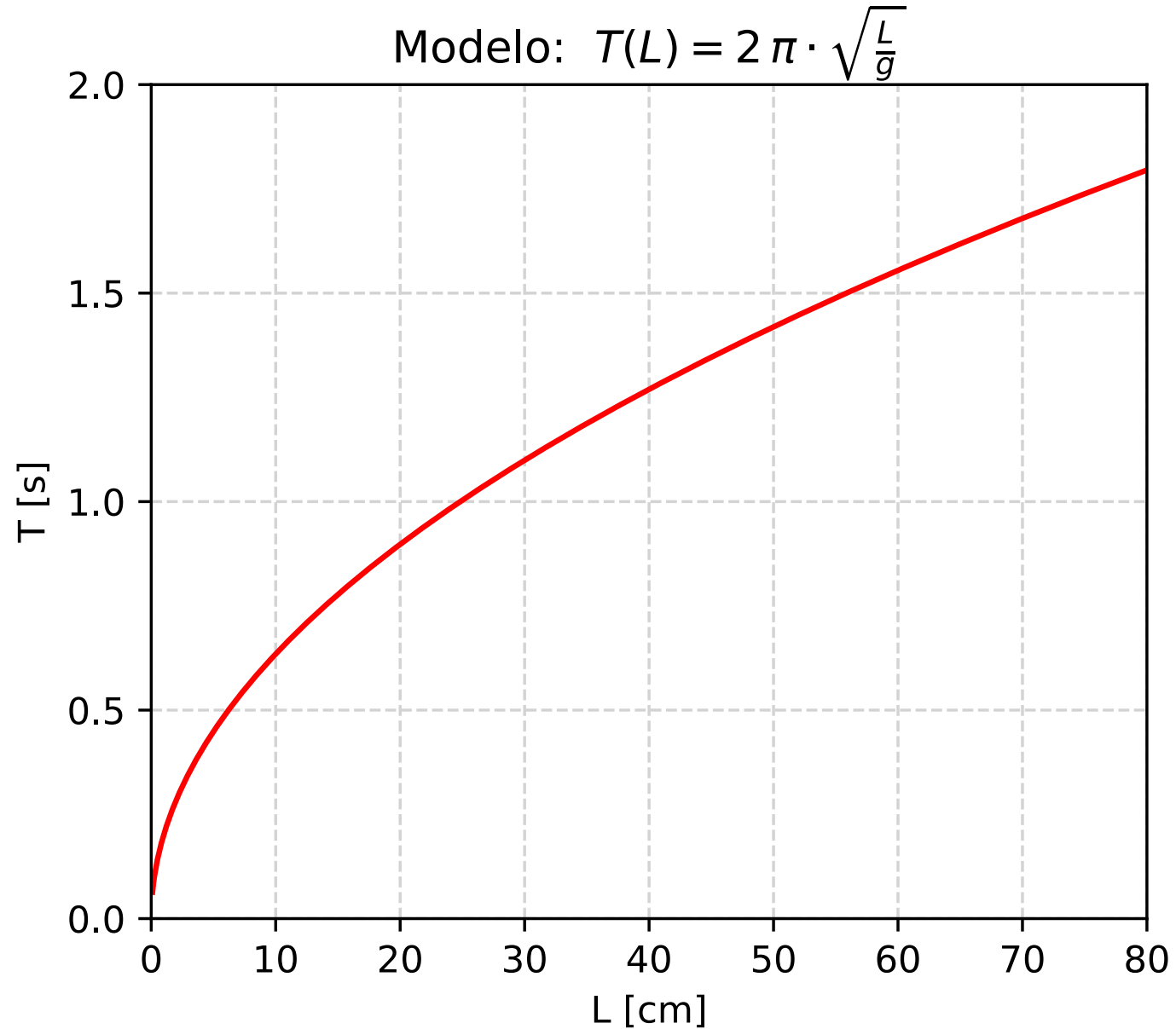
Modelo a validar:

$$T = 2\pi \sqrt{\frac{L}{g}}$$

Actividad: validar modelo péndulo

Modelo de péndulo

$$T = 2\pi \sqrt{\frac{L}{g}}$$



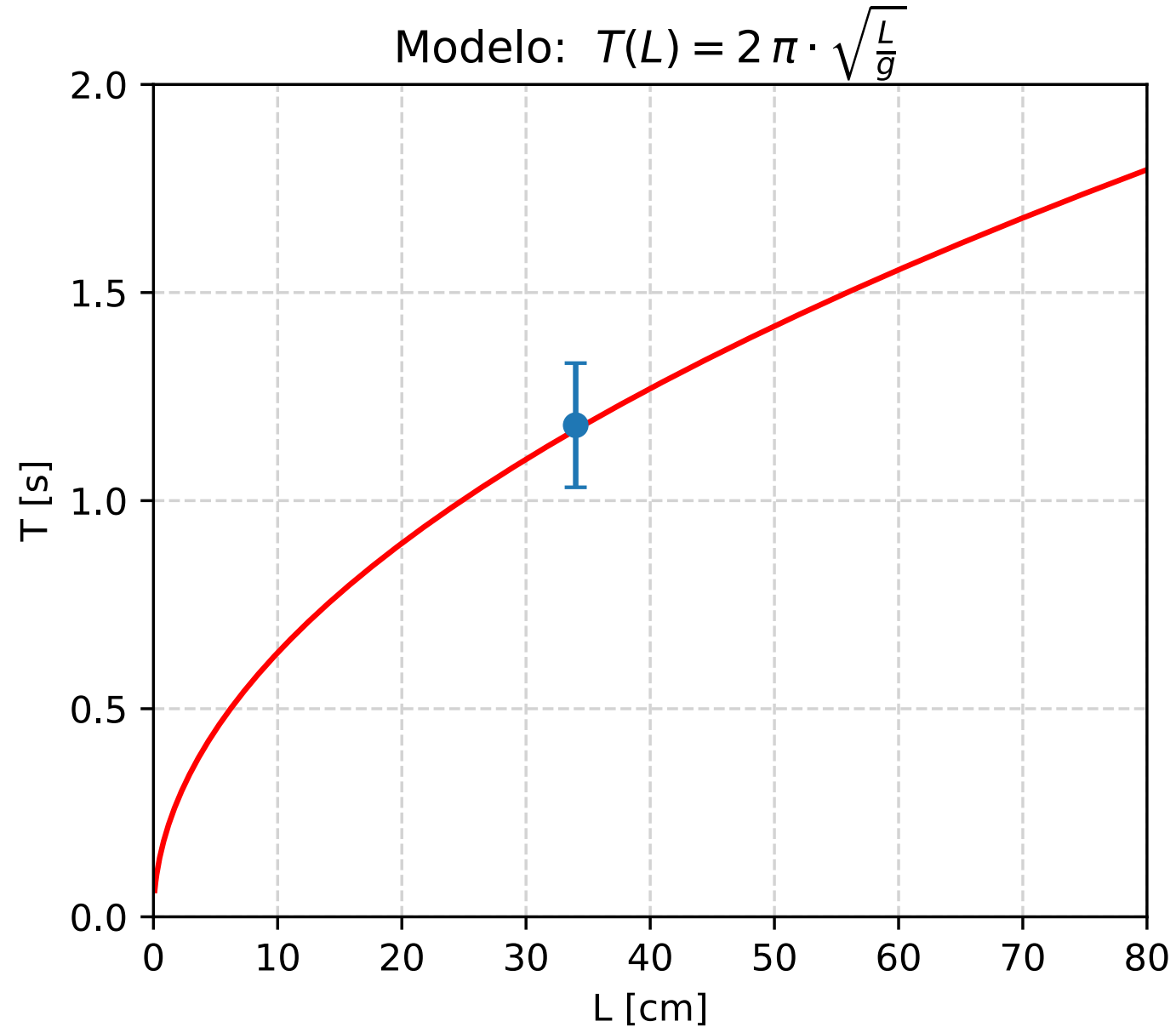
Actividad: validar modelo péndulo

Modelo de péndulo

$$T = 2\pi \sqrt{\frac{L}{g}}$$

La vez
pasada
tomamos
un dato

(T, L)



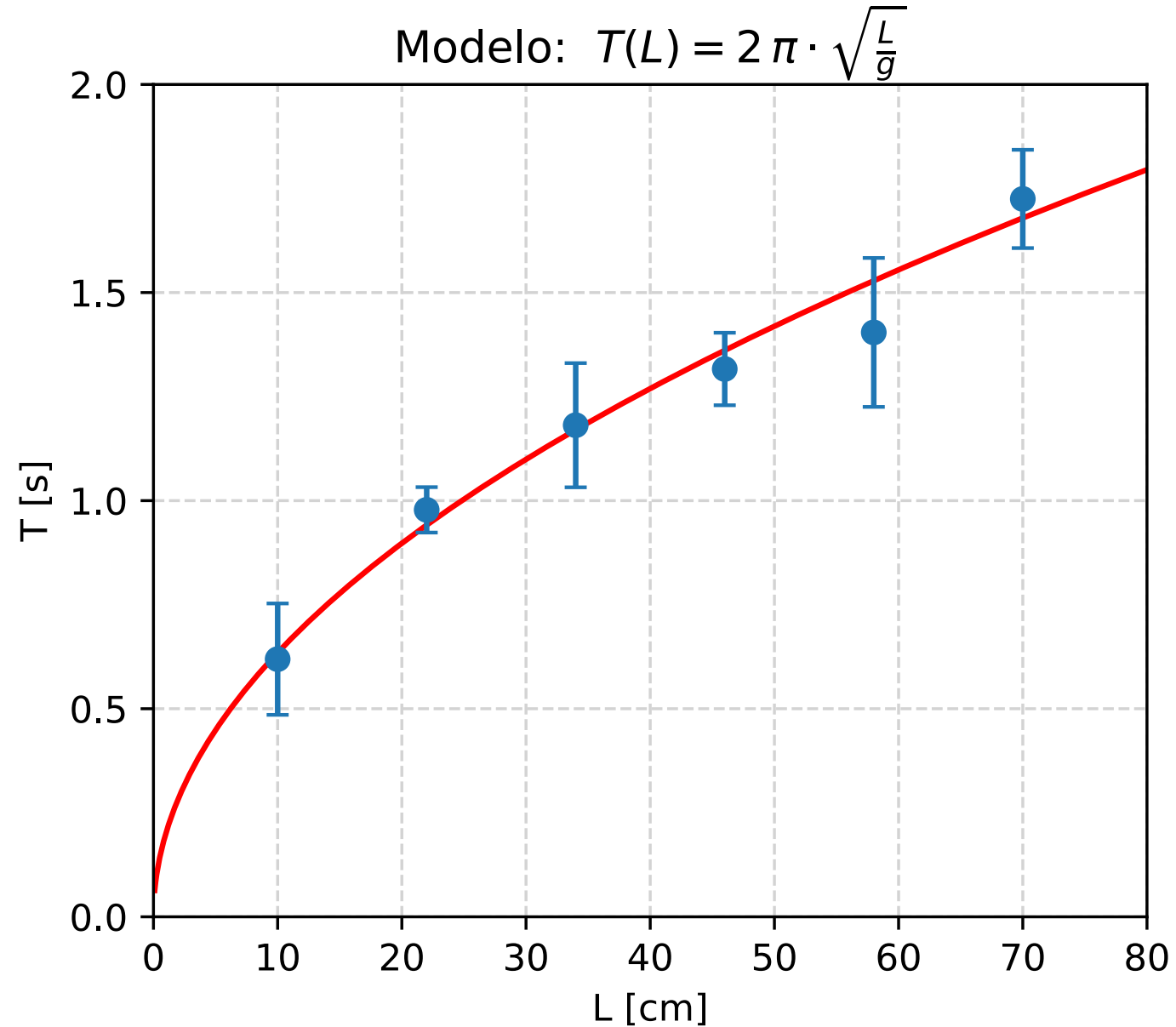
Actividad: validar modelo péndulo

Modelo de péndulo

$$T = 2\pi \sqrt{\frac{L}{g}}$$

Necesitamos
más pares
ordenados

$$(T_i, L_i)$$



RELOJES DE PÉNDULO

