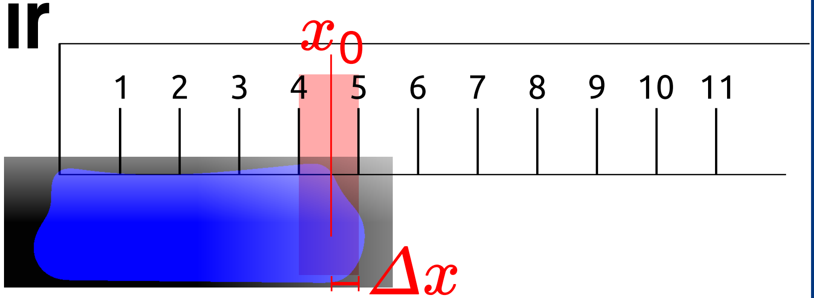
Clase 5

Comparación de datos con modelos (cuadrados mínimos)

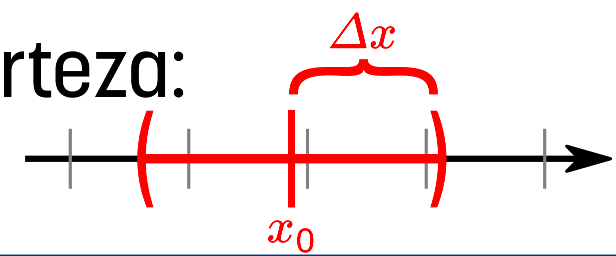
https://materias.df.uba.ar/l1d2021c2/

Repaso

Medir



con incerteza:

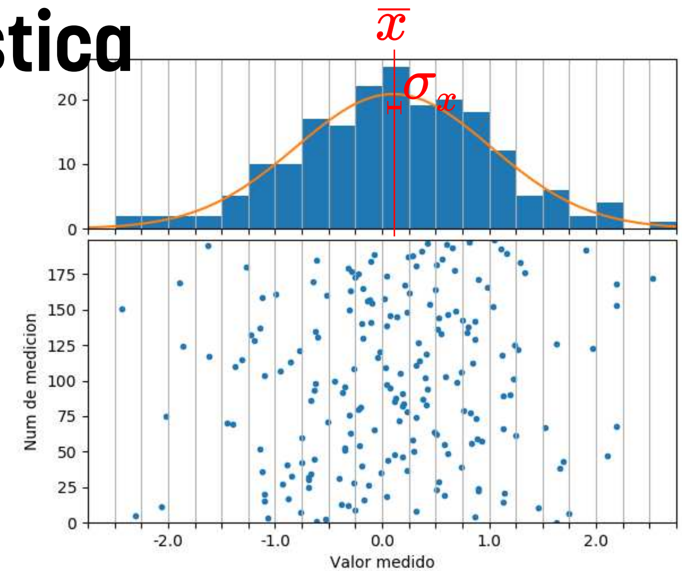


Estadística

$\overline{x}$

σ

σ_x



Calcular y propagar error

$$z = f(x, y) \quad (x_0, \Delta x) \quad (y_0, \Delta y)$$

$$z_0 = f(x_0, y_0) \quad \Delta z?$$

Medición INDIRECTA

¿Y para que queremos todo esto?

Repaso epistemológico

→ Las Ciencias Naturales son experimentales → **Corroboración de hipótesis**

→ El conocimiento científio es **EMPÍRICO** y **SISTEMÁTICO**

Ley física: Relaciones funcionales entre variables cuantitativas (magnitudes medibles)

→ Análisis de muchos datos

→ Comparación de datos con modelos

→ Corroboración de modelos

→ Predicciones

Repaso epistemológico

→ Las Ciencias Naturales son experimentales

→ El

Corroboración
de hipótesis

L
Re
fu

V
cu
(magnitudes
medibles)

Digresión

→ Corroboración de modelos
→ Predicciones

Comparación de datos y modelos

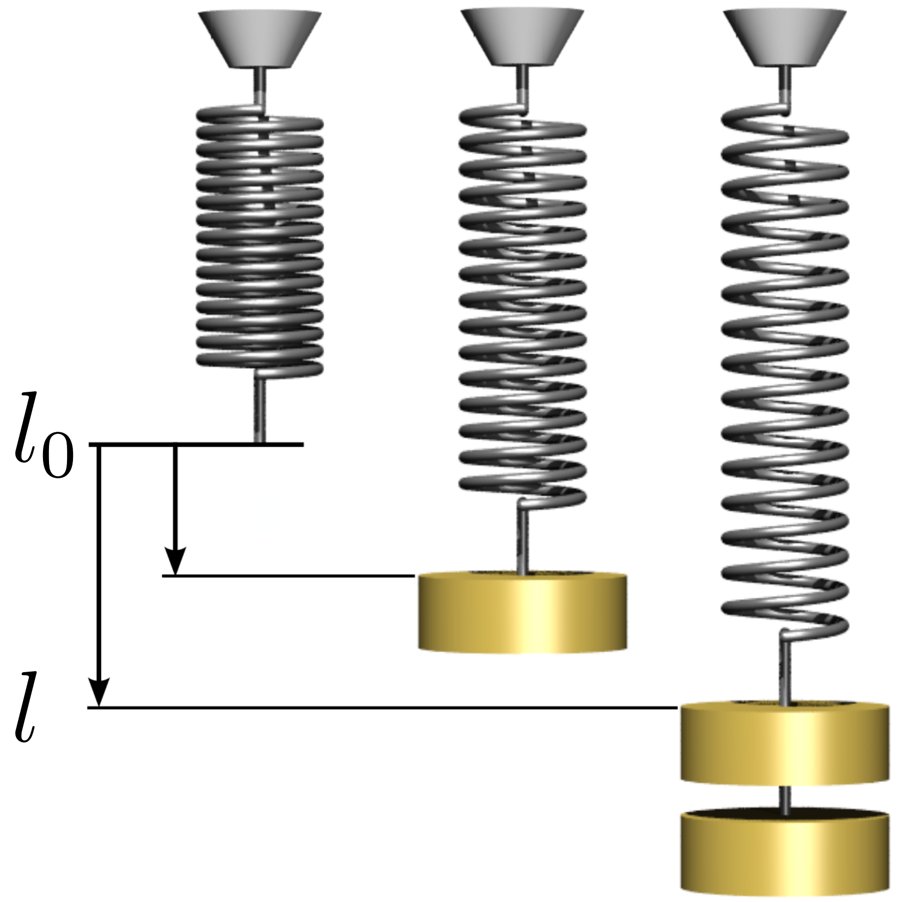
Ejemplo de modelo: Ley de Hook

Relación funcional entre posición y fuerza: Lineal

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

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

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



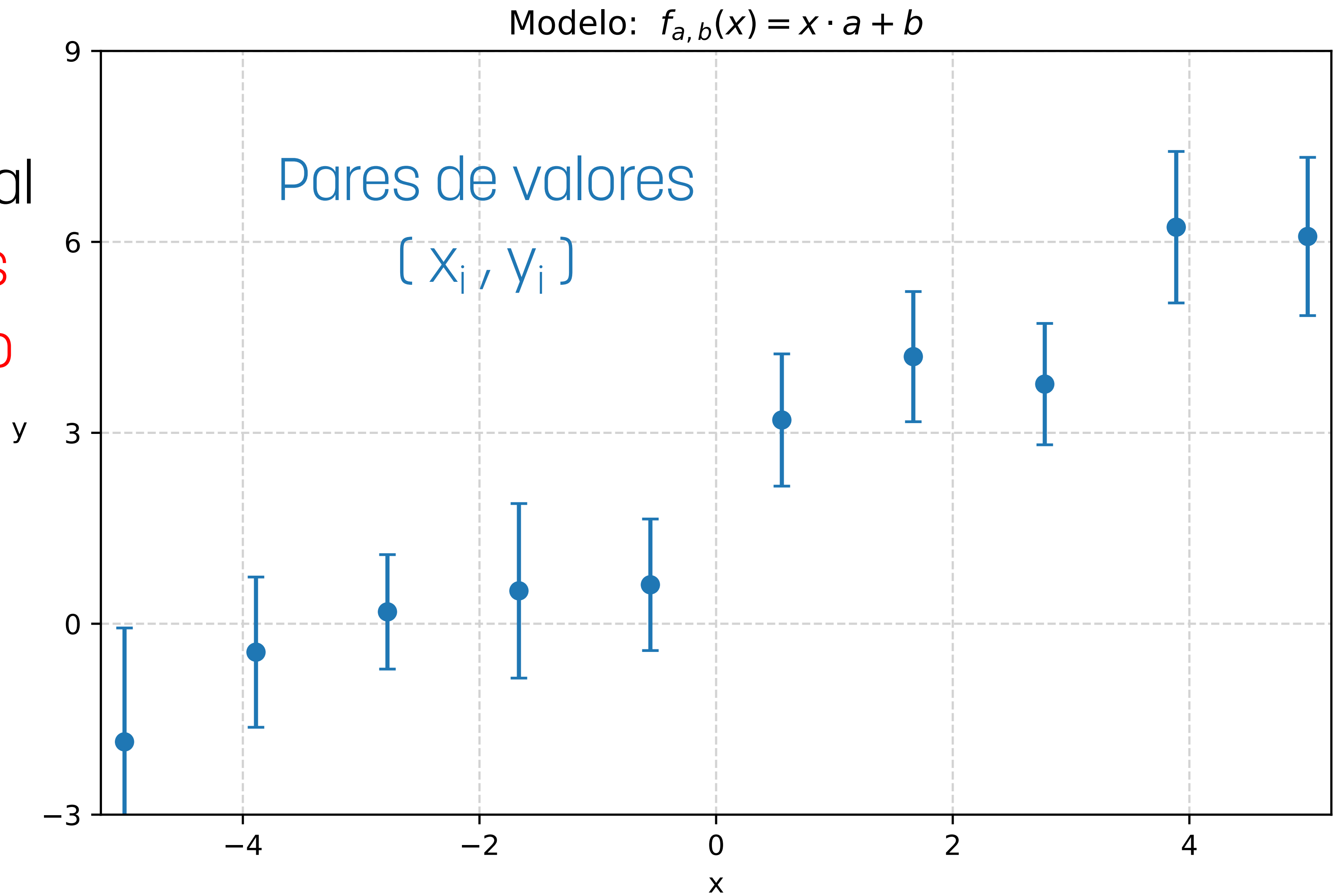
Cuadrados mínimos

Conjunto de datos medidos (con sus errores)

Modelo Lineal

parámetros

f(x) = x·a + b

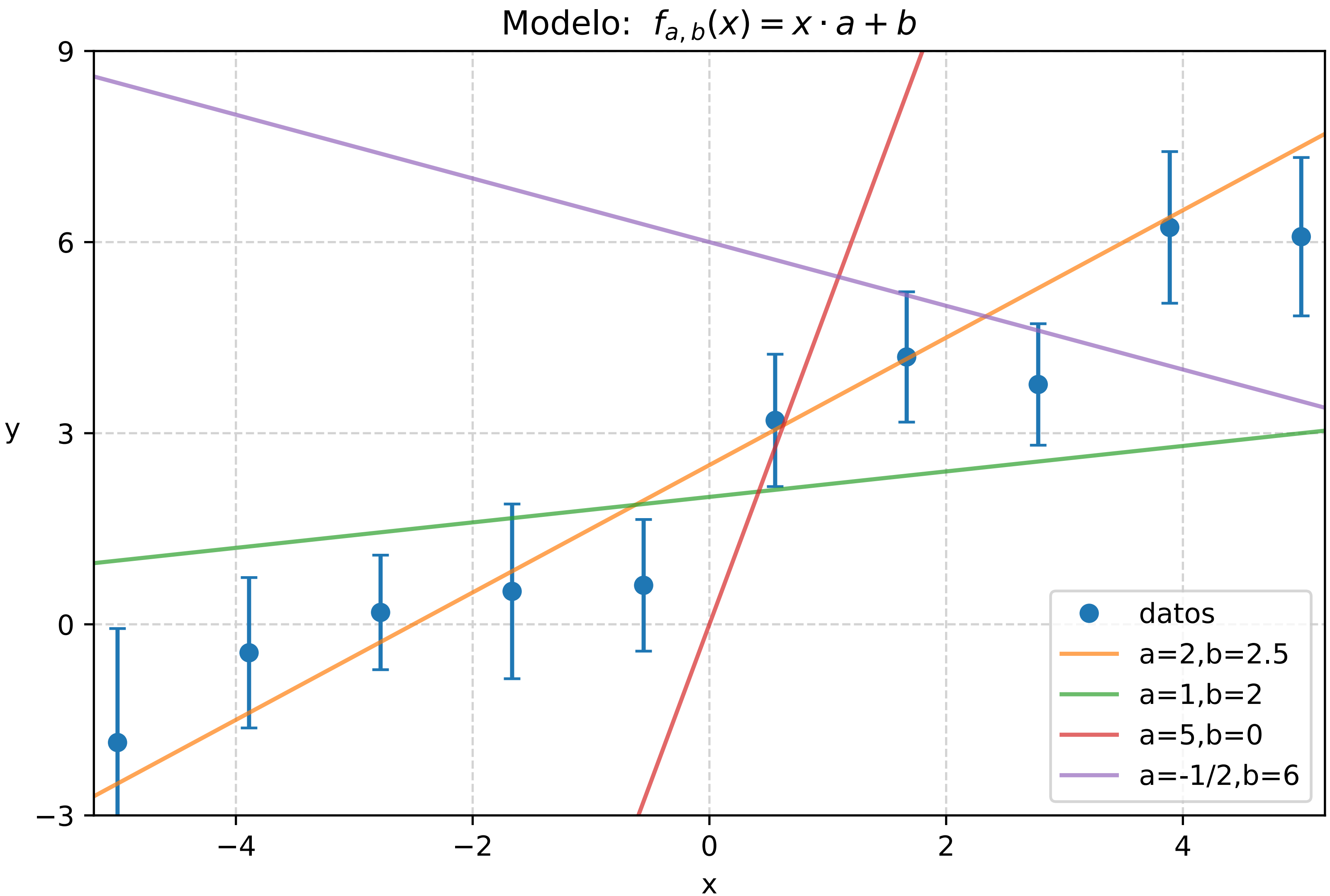


¿Responden los datos al modelo?

Cuadrados mínimos

Modelo para diferentes parámetros

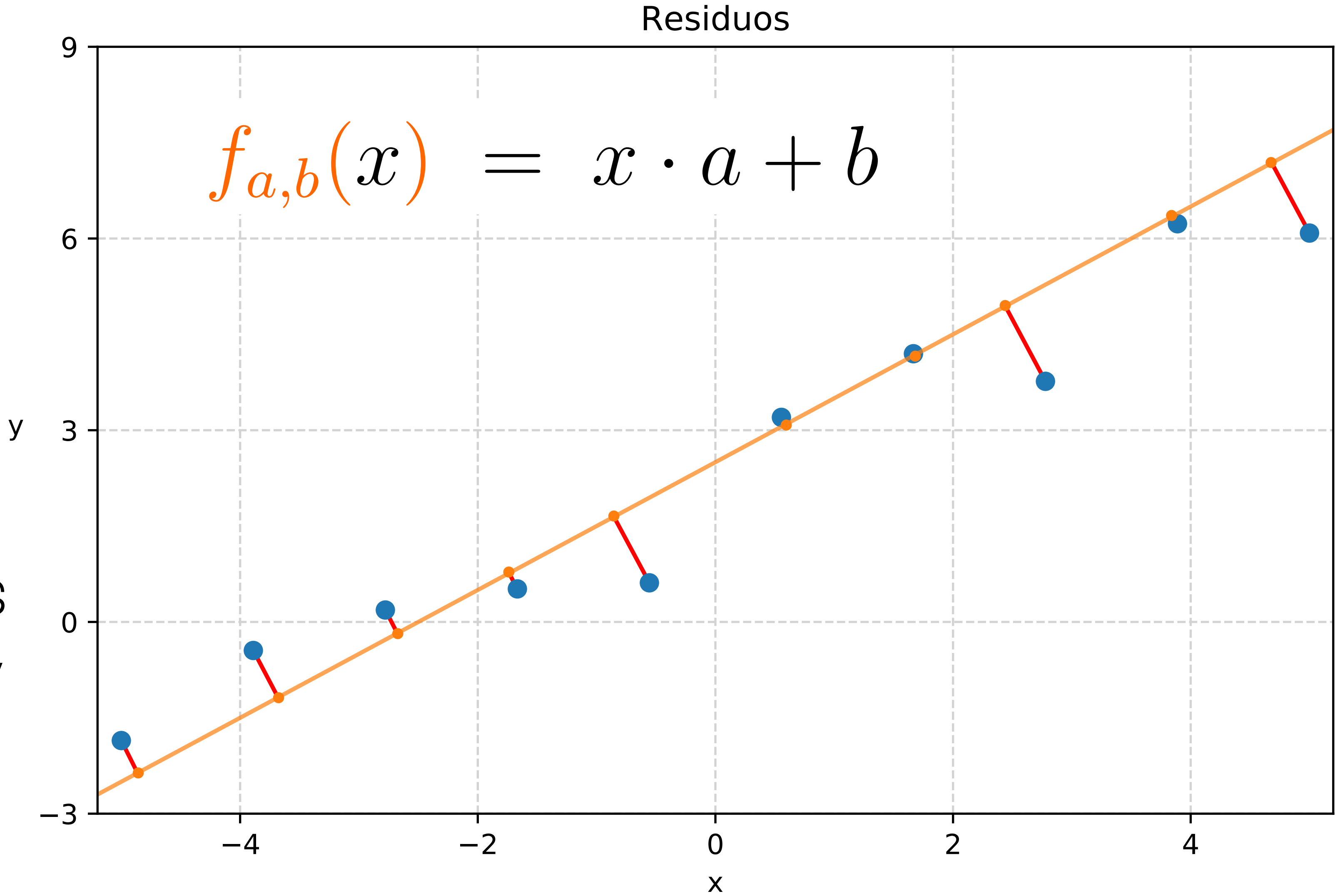
¿Con que parámetros el modelo *"se parece"* más a los datos?



Cuadrados mínimos

$$(x_i, y_i)$$

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



$$f_{a,b}(x) \ = \ x \cdot a + b$$

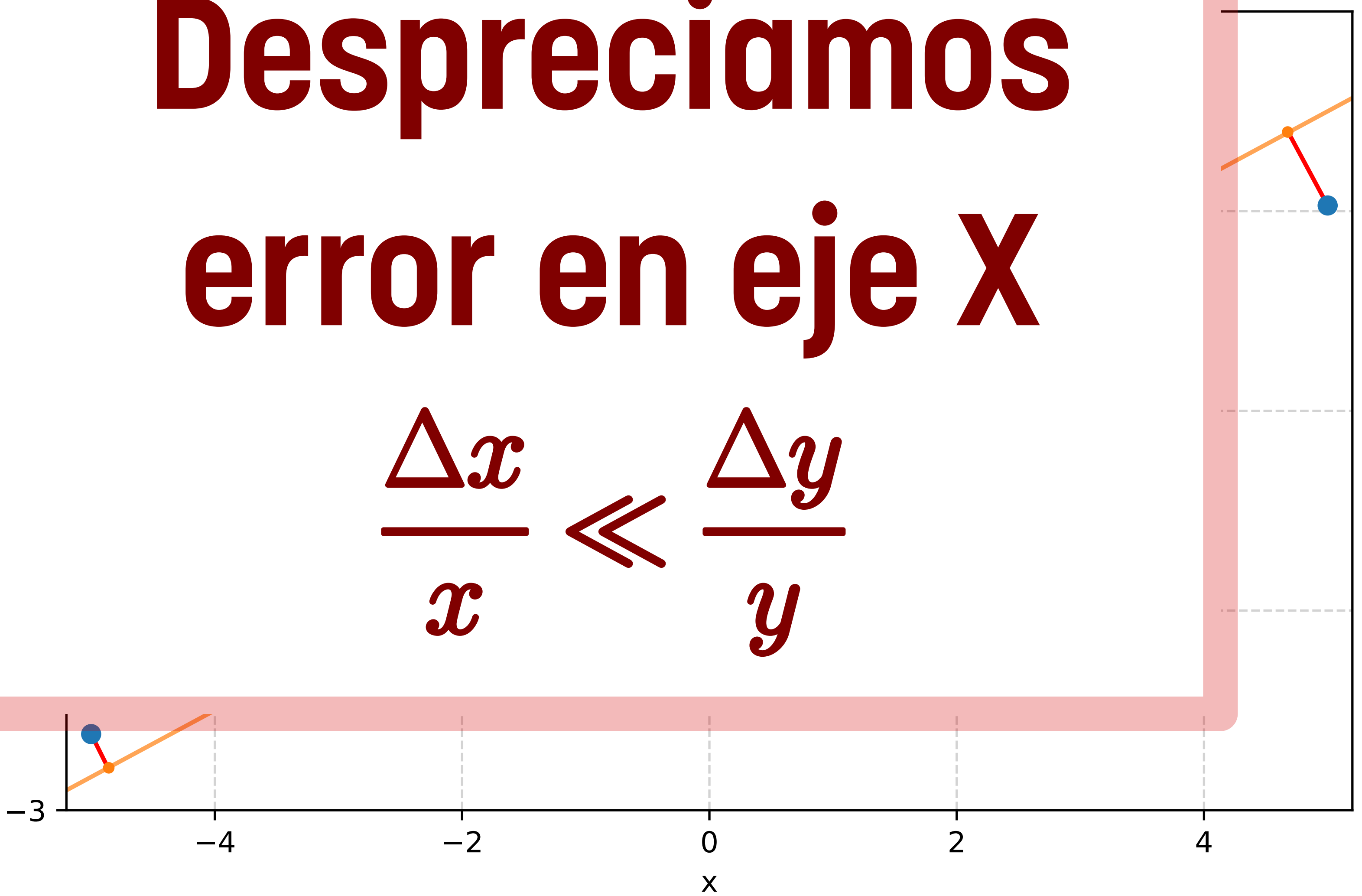
Cuadrados mínimos

$$(x_i, y_i)$$

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

Despreciamos error en eje X

$$\frac{\Delta x}{x} \ll \frac{\Delta y}{y}$$



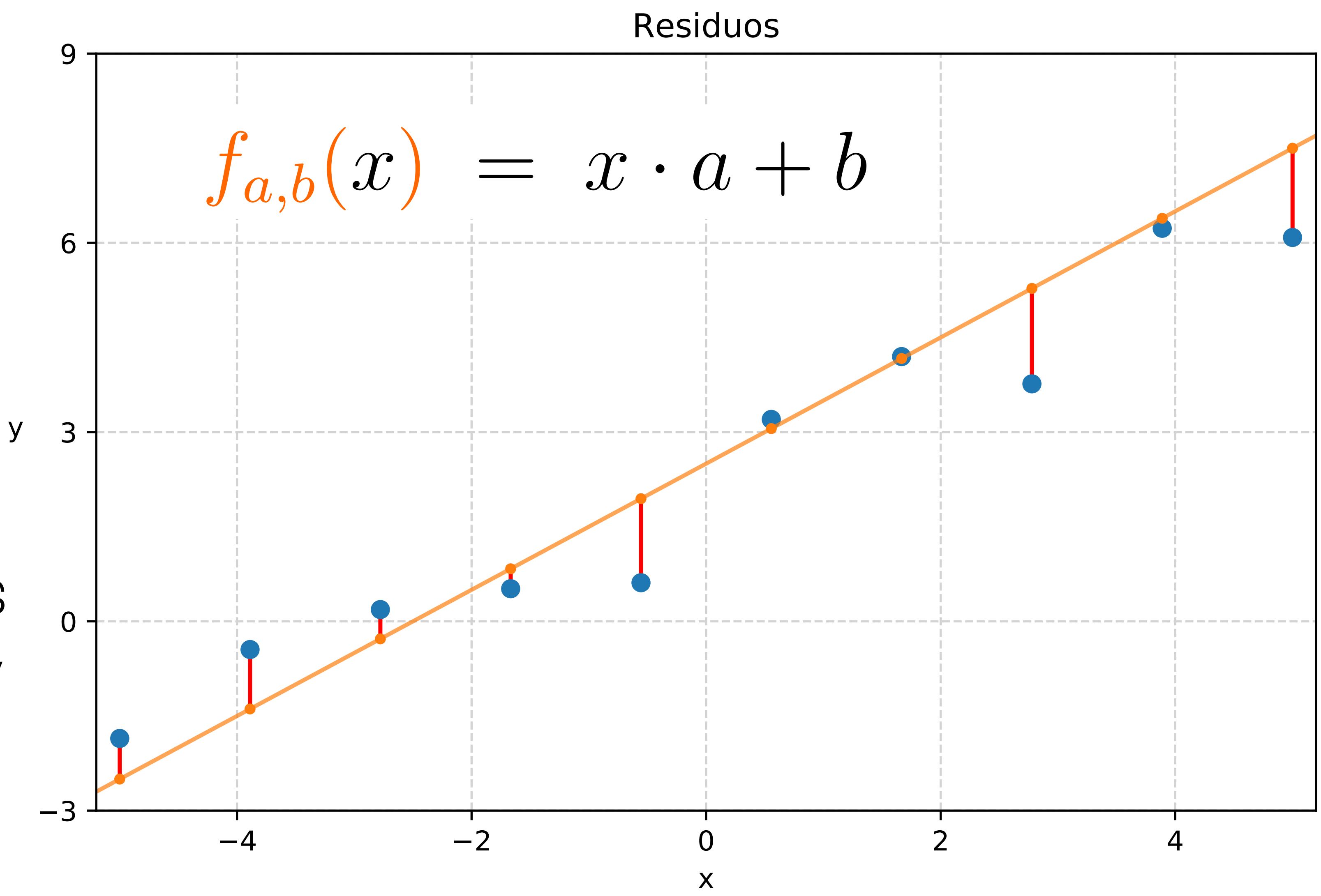
Cuadrados mínimos

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

$$(x_i, y_i)$$

$$f_{a,b}(x) = x \cdot a + b$$

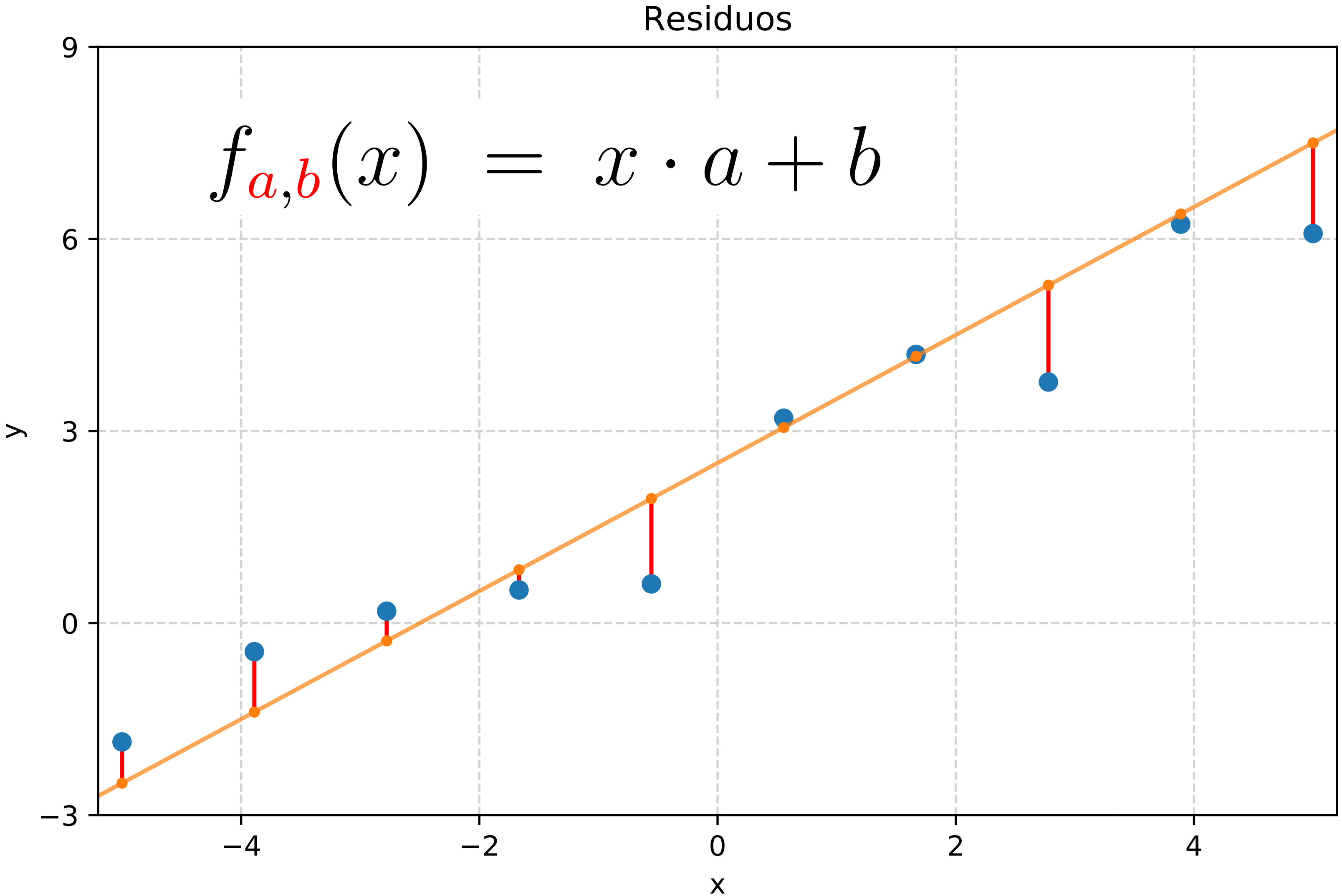
Fabricar una *"medida de distancia"* entre las predicciones del modelo y los datos medidos



Cuadrados mínimos

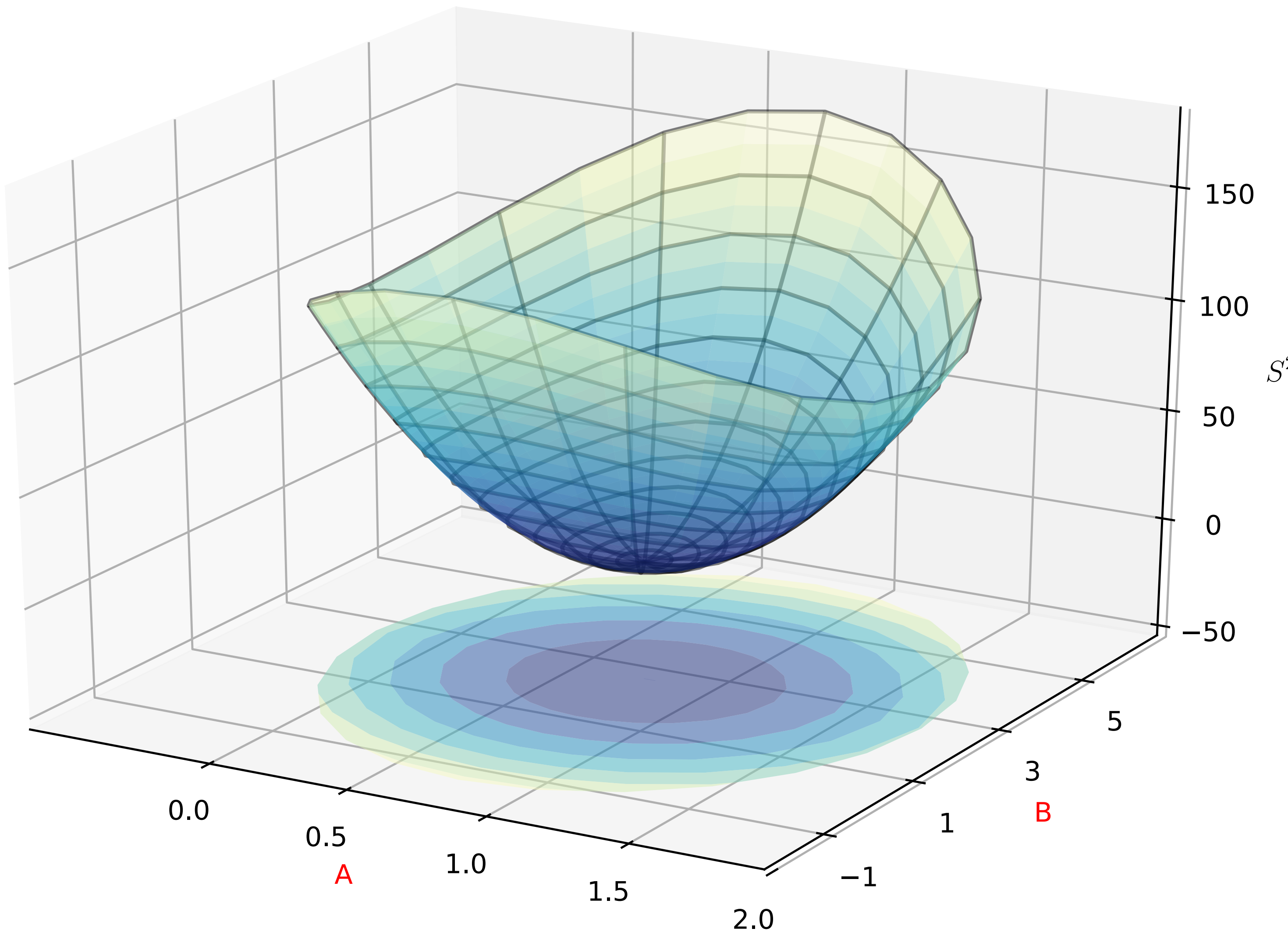
$$S(a,b) = \sum_{i=1}^{N} [y_i - f_{a,b}(x_i)]^2 = \sum_{i=1}^{N} [y_i - x_i \cdot a - b]^2$$

$$(x_i, y_i)$$



Cuadrados mínimos

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



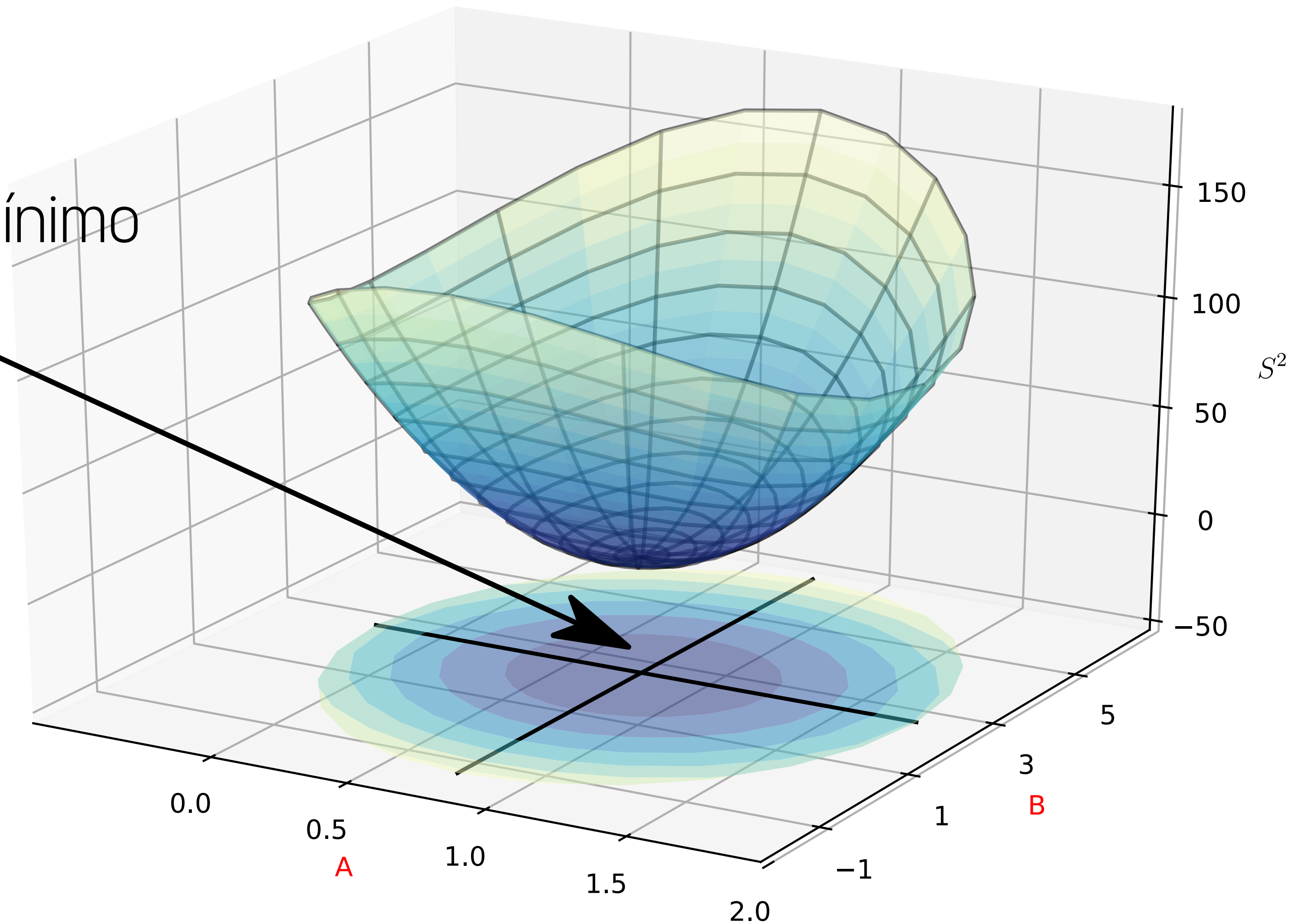
Superficie que representa a S^2

Cuadrados mínimos

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

Búsqueda del mínimo

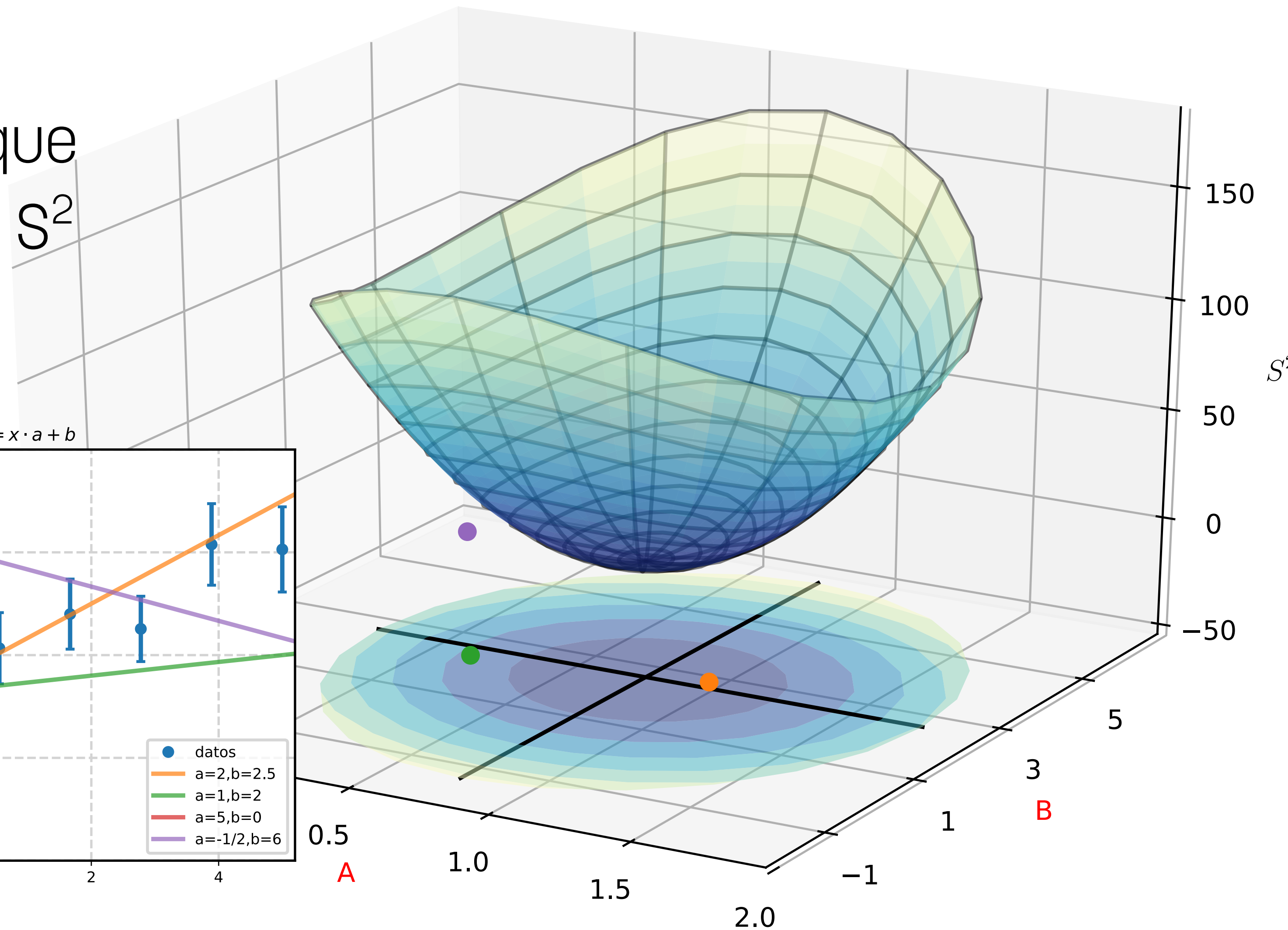
Algoritmo

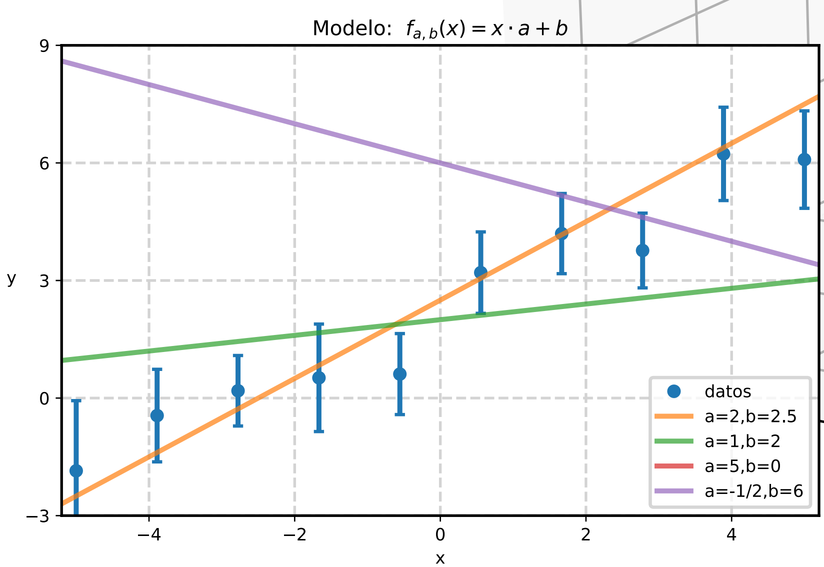


Cuadrados mínimos

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

Ejemplos de parámetros que no minimizan S^2





Cuadrados mínimos

Algoritmo para modelo lineal (sin incertezas en y)

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

$$\left.\begin{aligned} \frac{\partial S}{\partial a} &= 0 \\ \frac{\partial S}{\partial b} &= 0 \end{aligned}\right\}$$

Sistema de 2 ecuaciones lineales con 2 incógnitas

Cuadrados mínimos

Algoritmo para modelo lineal (sin incertezas en y)

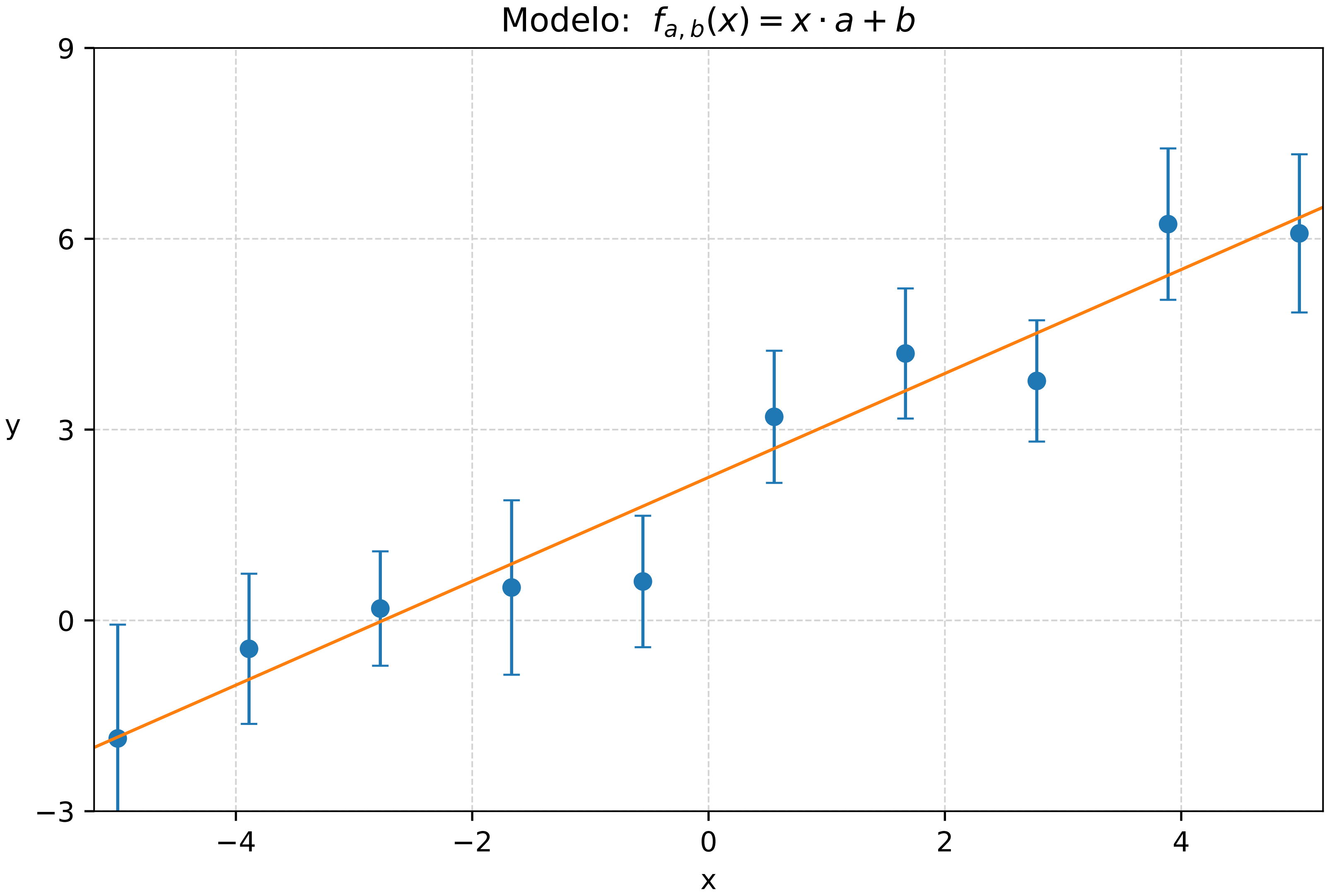
$$S\ (a, b) = \sum_{i=1}^{N} \left[y_i - x_i \cdot a - b \right]^2$$

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

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

Cuadrados mínimos

Con los parámetros que minimizan S obtenemos el modelo ajustado



Cuadrados mínimos

Algoritmo para modelo lineal (con incertezas en y)

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

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

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

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

Cuadrados mínimos

Algoritmo para modelo lineal (con incertezas en y)

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

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

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

$$\Delta = \sum \frac{1}{\sigma_{y_i}^2} \sum \frac{x_i^2}{\sigma_{y_i}^2} \cdots$$

Parámetros y calidad de ajuste

- → **¿Que tan bueno es mi ajuste?**
  - → ¿Corresponde ese modelo al fenómeno bajo estas hipótesis?
  - → ¿Entre dos o mas modelos… ¿cual se corresponde más?
  - → ¿Hay contenido físico que no estoy contemplando?

- → **¿Cuanto se sobre los parámetros inferidos?**
  - → ¿Son independientes?
  - → Si uso el modelo como medición indirecta:
    ¿Cual es la incerteza de mi medición?

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

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

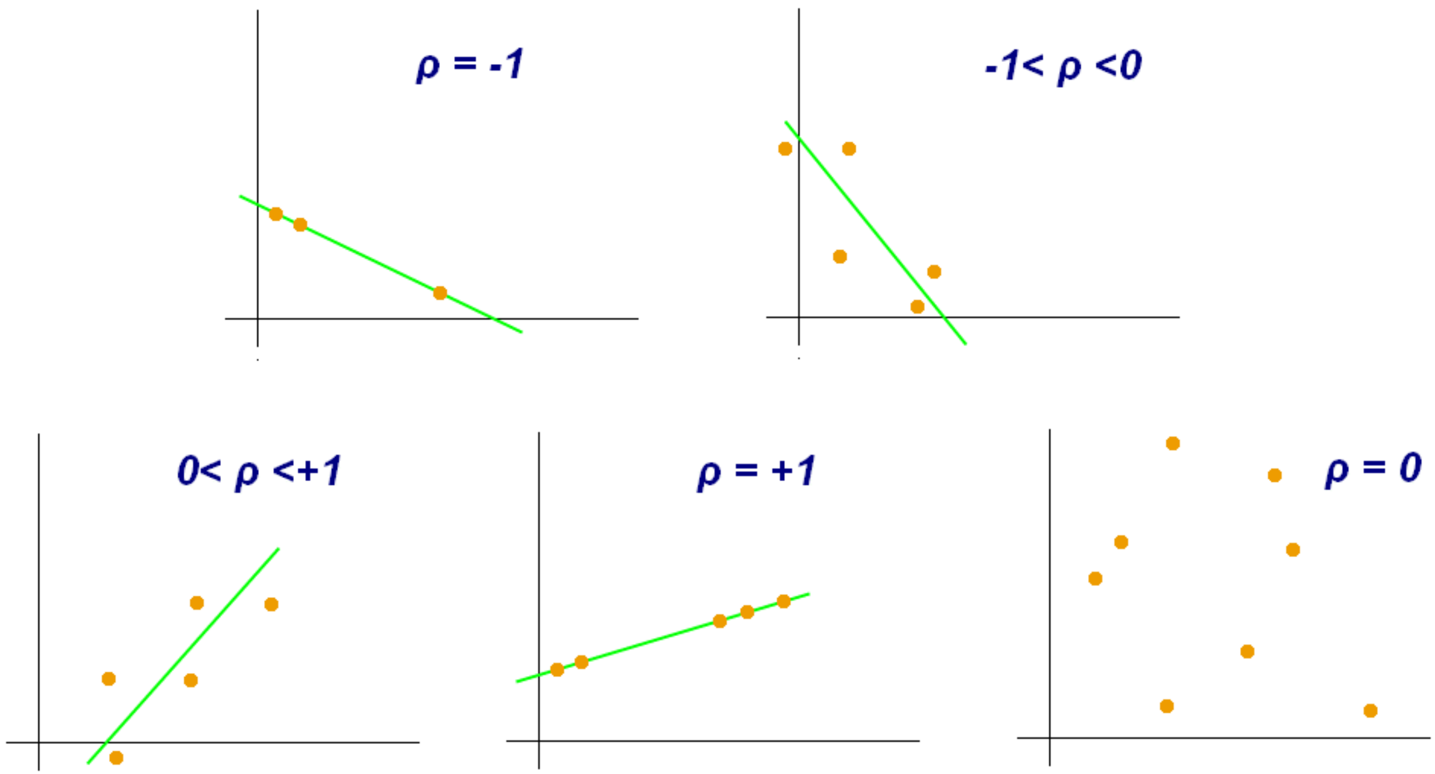
$$\mathrm{RSS} \equiv S = \sum_i \mathrm{res}_i^2$$

$$\mathrm{SST} \equiv \sum_i (y_i - < y >)^2$$

$$R^2_{adj} = 1 - \frac{\mathrm{RSS}}{\mathrm{SST}} \cdot \frac{N-1}{N-P-1}$$

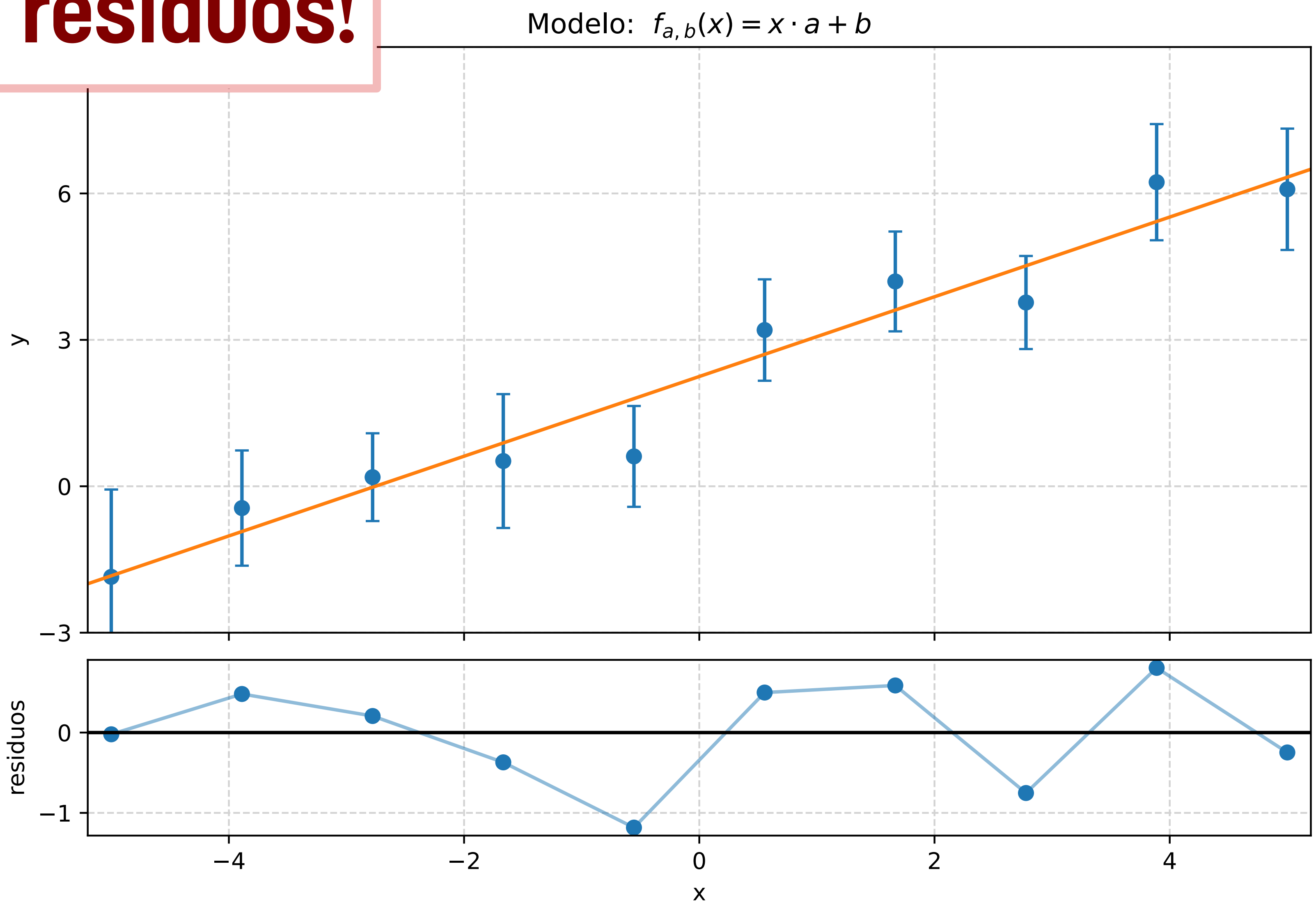
Estimadores de bondad de ajuste

- **Coeficiente de Correlación de Pearson, r , r-pear , ρ**
  - "Cuando una variable sube la otra también" o al revez
  - $-1 < \rho < 1$
  - $|\rho| = 1$ significa que los datos estan perfectamente alineados

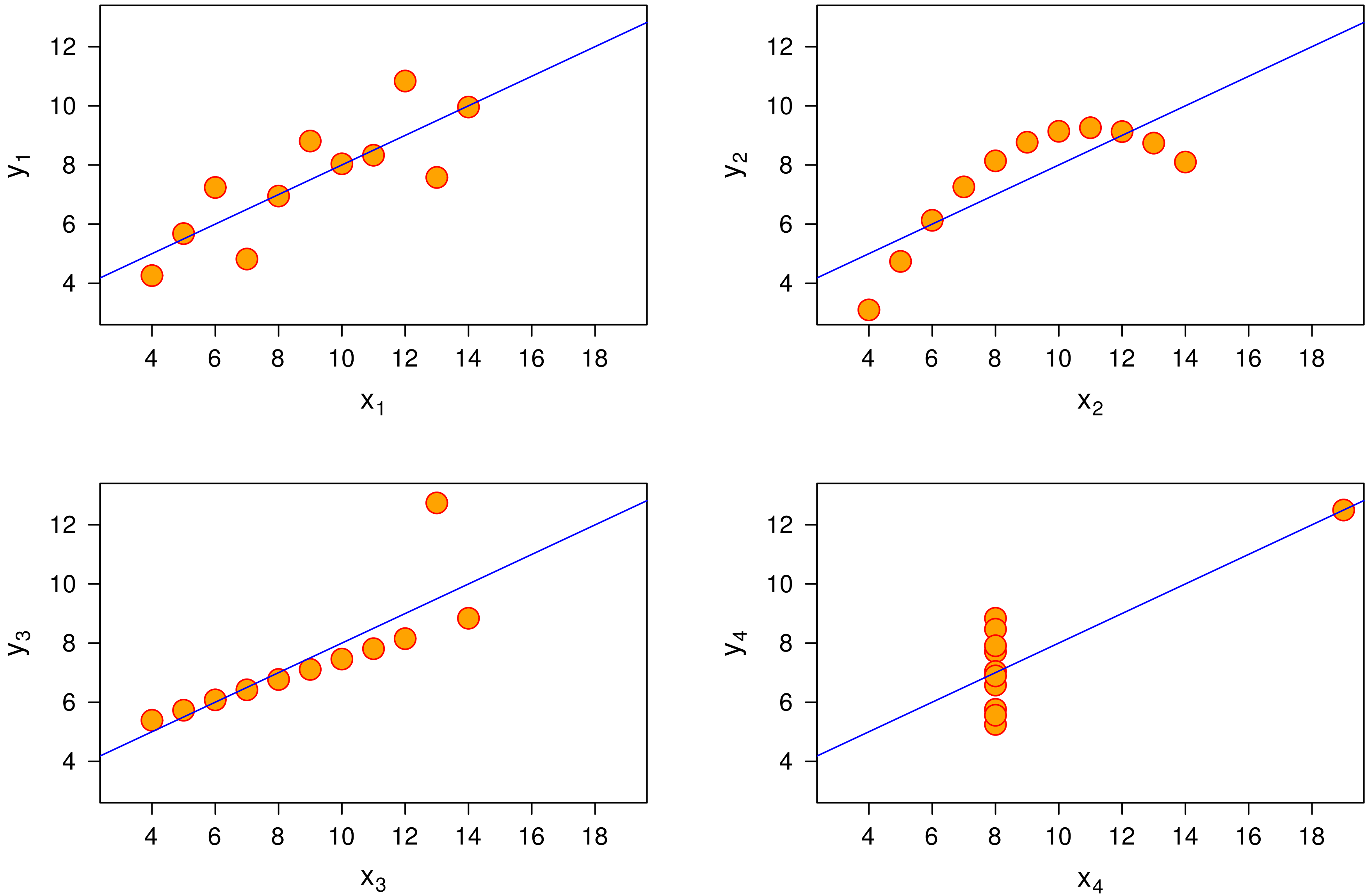


Analizar residuos

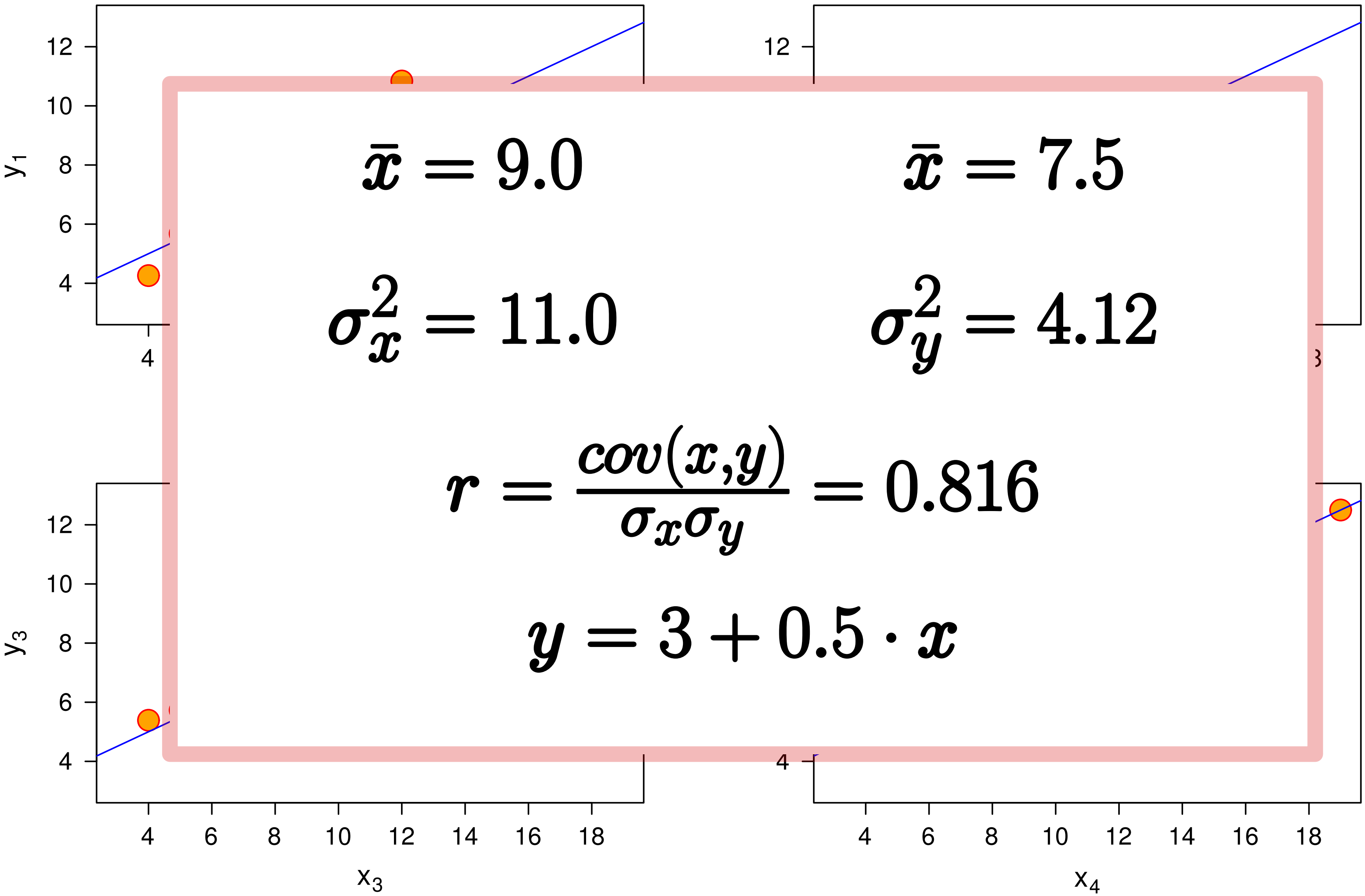
Importante: analizar residuos!



Cuarteto de Anscombe



Cuarteto de Anscombe



$$\bar{x} = 9.0 \qquad \bar{x} = 7.5$$

$$\sigma_x^2 = 11.0 \qquad \sigma_y^2 = 4.12$$

$$r = \frac{cov(x,y)}{\sigma_x \sigma_y} = 0.816$$

$$y = 3 + 0.5 \cdot x$$

¿...y si el modelo no es lineal?

Tácticas para ajustar modelos que no son lineales

¿...y si el modelo no es lineal?

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

Ej: Caida libre por aceleración de la gravedad

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

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

$$\boxed{Y = X \cdot a}$$

$$a \equiv \frac{1}{2}g$$

$$X \equiv t^2$$

¿...y si el modelo no es lineal?

→ **Usar cantidades transformadas (ej: aplicar logaritmo)**

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

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

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

$$\boxed{Y = a \cdot X + b}$$

$$a \equiv 2$$

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

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

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

¿...y si el modelo no es lineal?

→ **Usar cantidades transformadas (ej: aplicar logaritmo)**

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

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

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

$$\boxed{Y = a \cdot X + b}$$

$$a \equiv 2$$

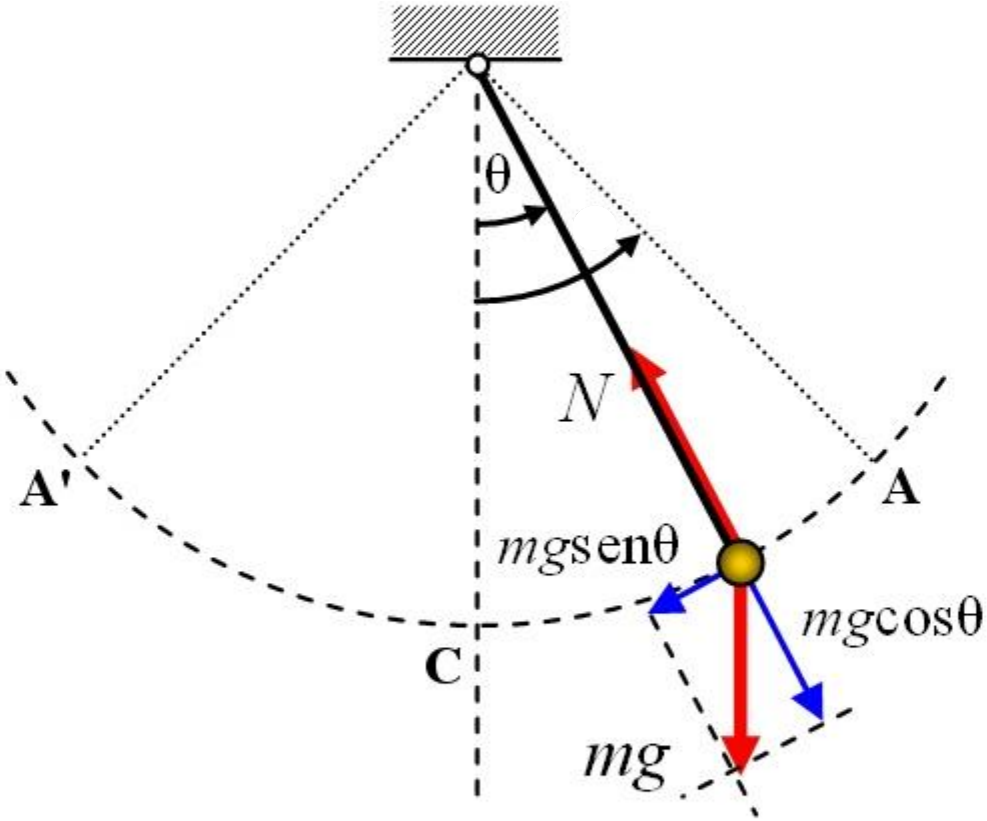
$$b \equiv \log\left(\frac{1}{2}g\right)$$

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

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

Actividad: 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: $\sin(\theta) \sim \theta \qquad \theta \ll 1$

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

Solución:

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

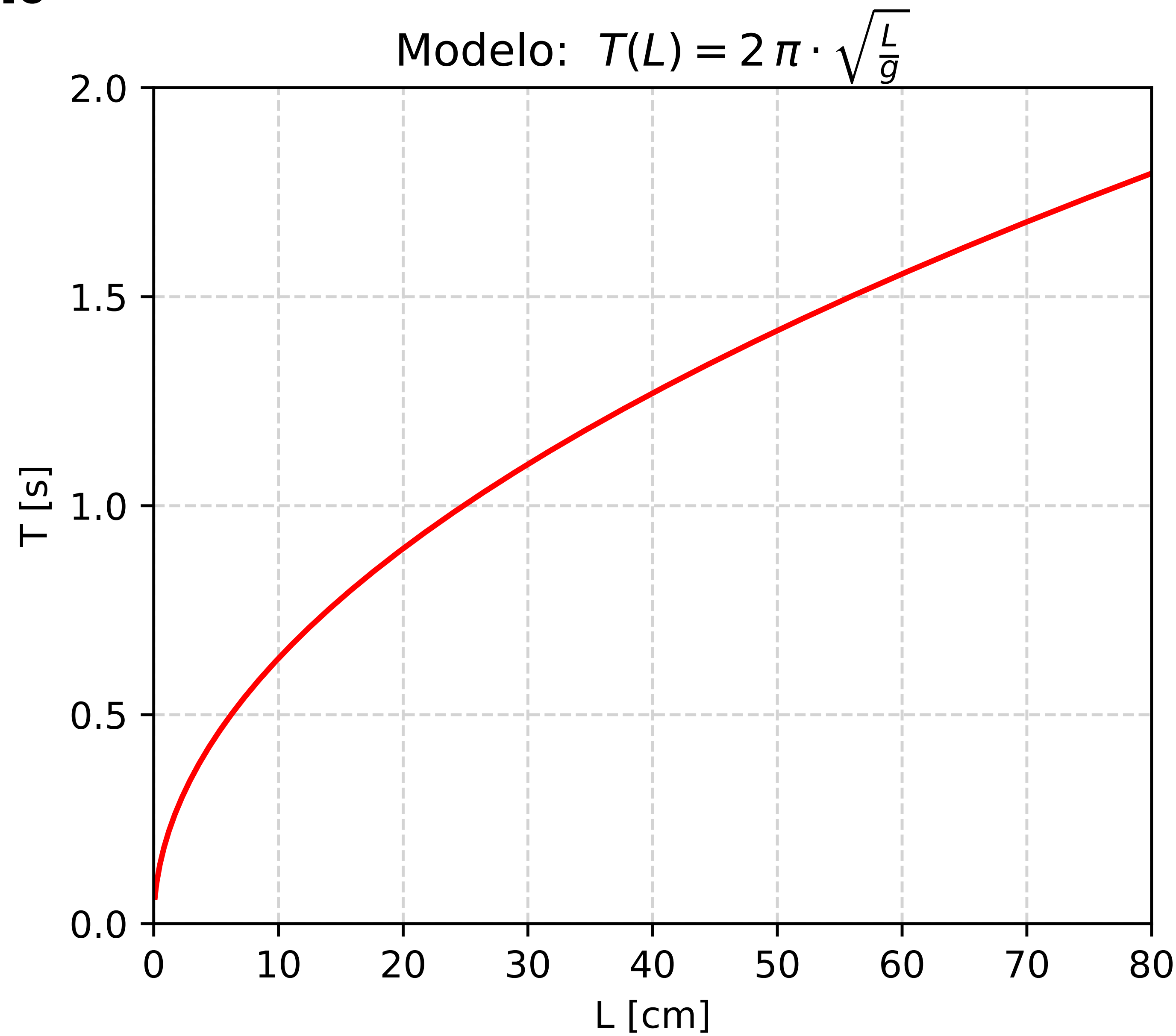
$$\omega = \frac{2\,\pi}{T} = \sqrt{\frac{g}{L}}$$

Modelo a validar: $\boxed{T = 2\pi \sqrt{\frac{L}{g}}}$

Actividad: validar modelo péndulo

Modelo de péndulo

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



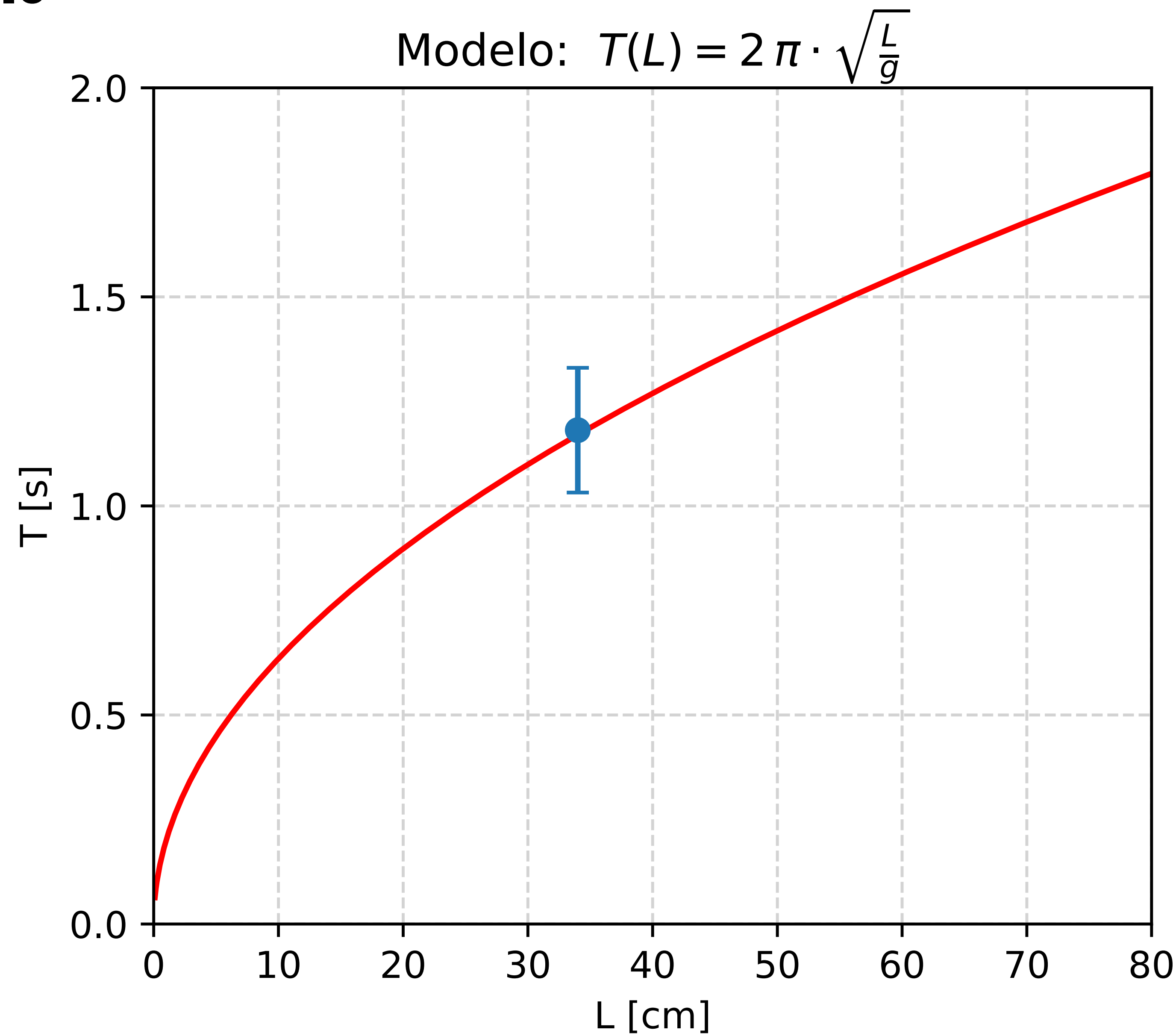
Actividad: validar modelo péndulo

Modelo de péndulo

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

La vez pasada tomamos un dato

$$(T , L)$$



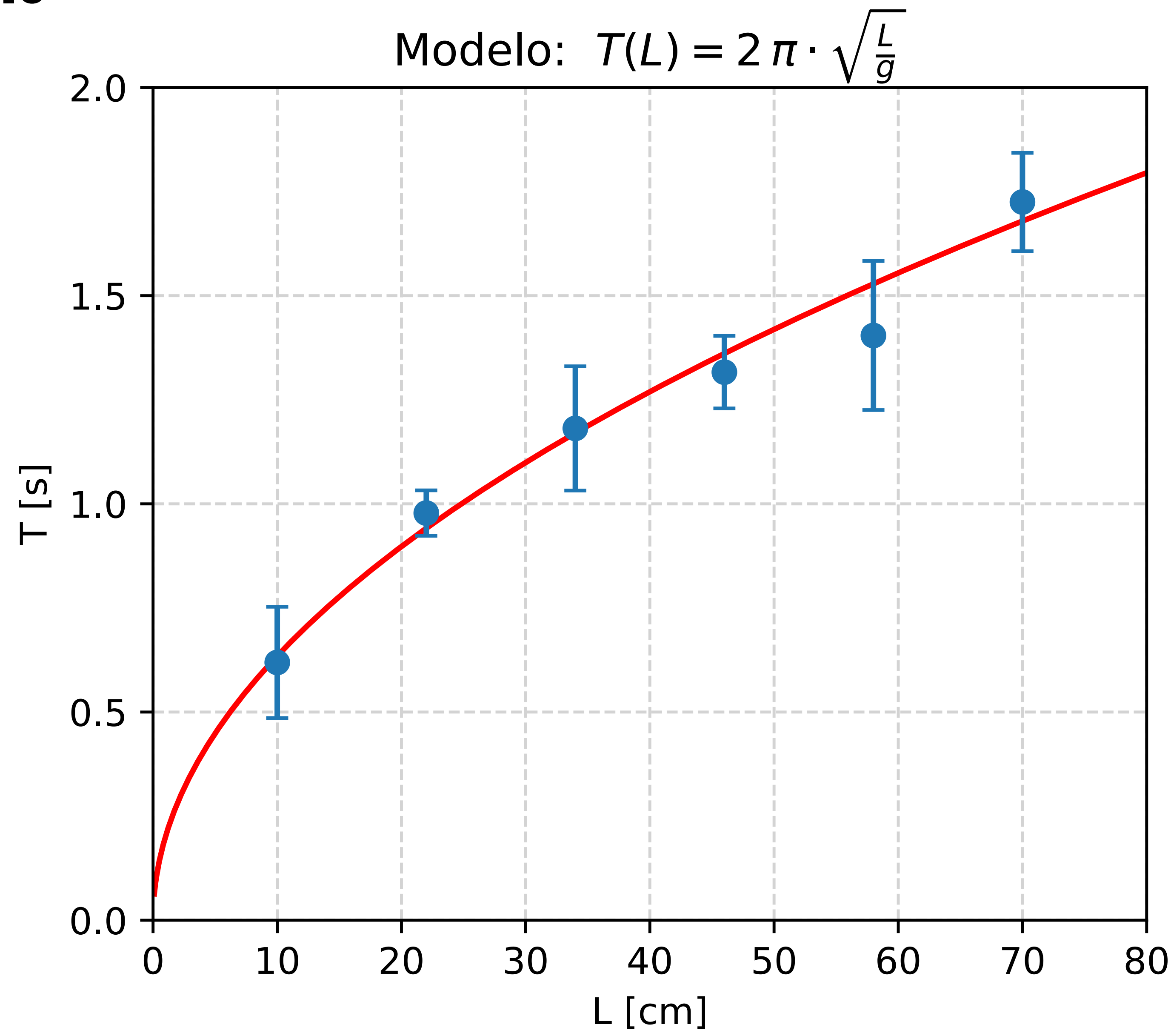
Actividad: validar modelo péndulo

Modelo de péndulo

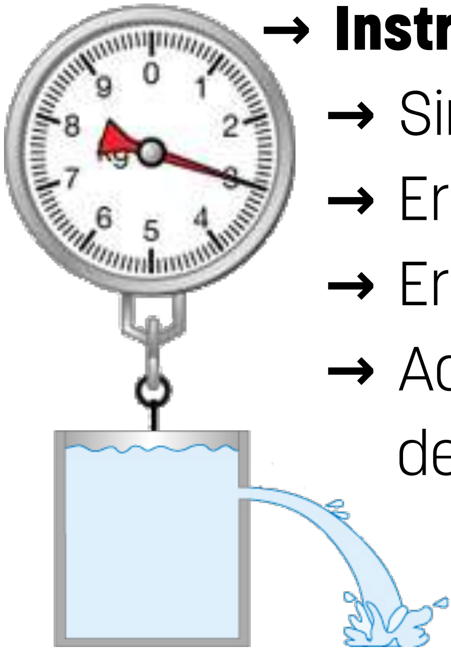
$$T = 2\pi \sqrt{\frac{L}{g}}$$

Necesitamos más pares ordenados

$$(T_i, L_i)$$

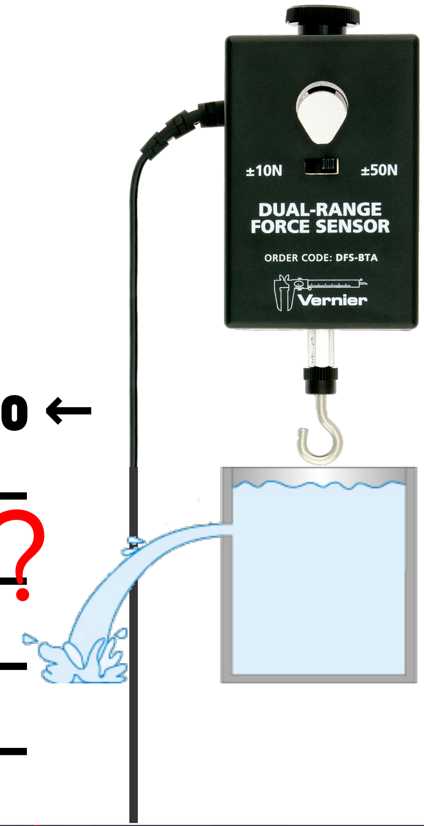


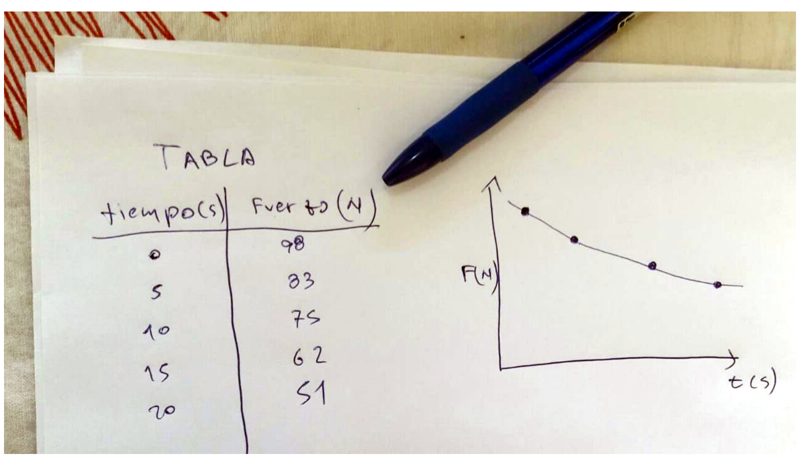
Instrumentación

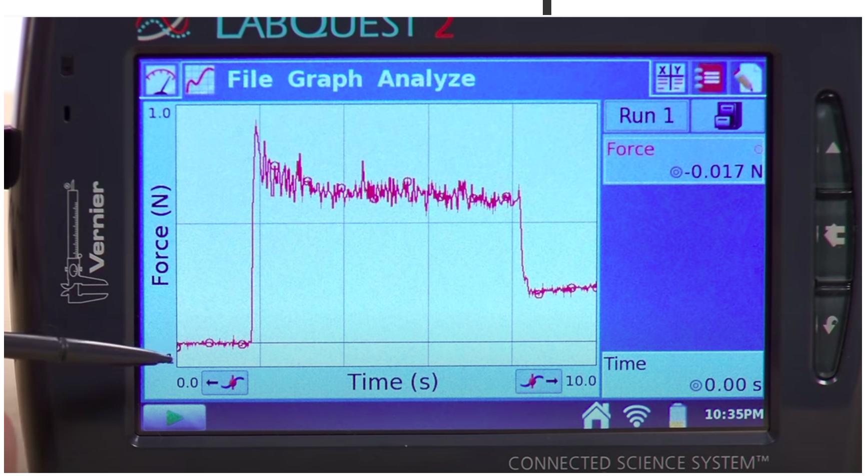


→ **Instrumento pasivo**
- → Simple
- → Error Humano
- → Error apreciación
- → Adquisición manual de pocos datos

Instrumento activo ←
- Más complejo ←
- ~~Error Humano~~ ← ?
- Error apreciación menor ←
- Adquisición masiva ←



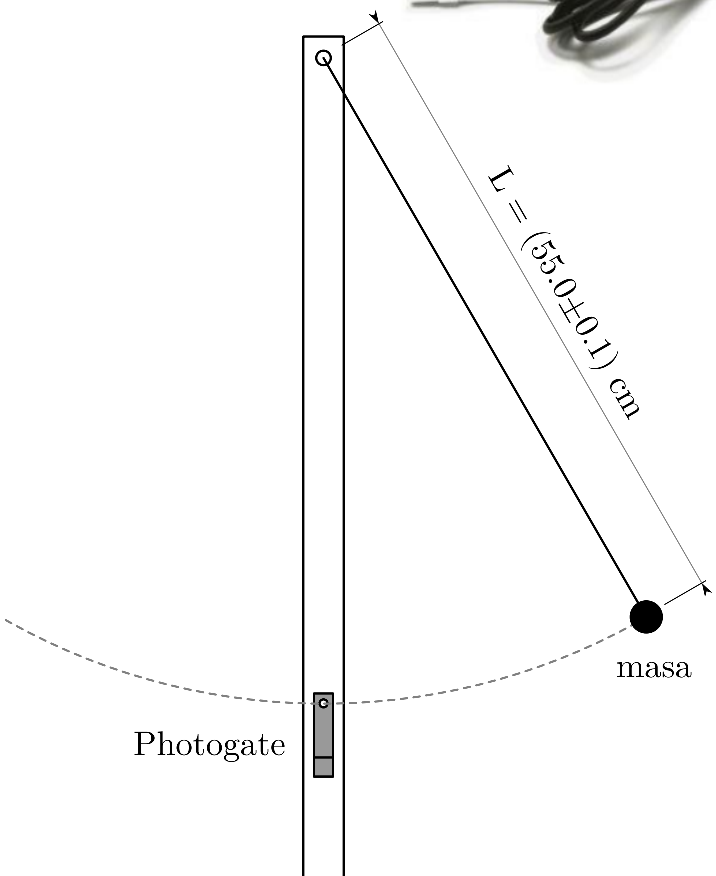


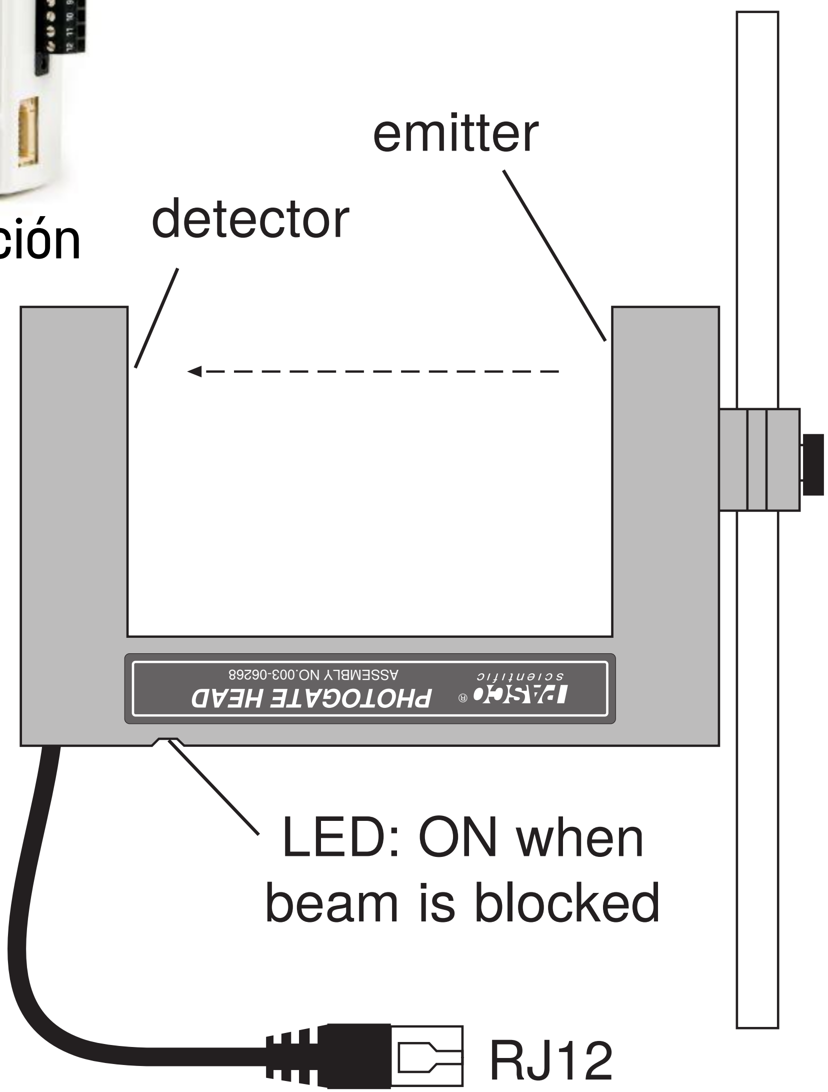


Periodo con "Photogate"

Fotosensor y led infrarrojo

Digitalización





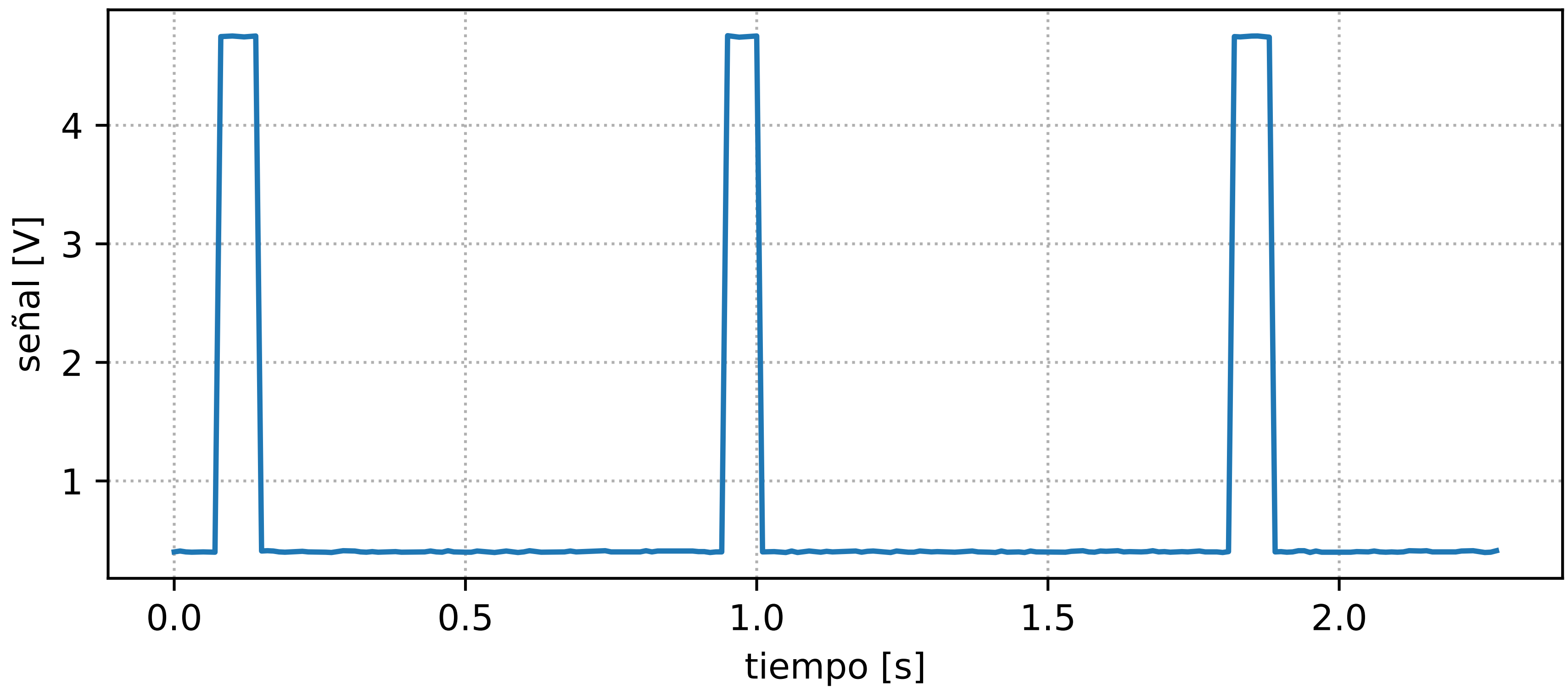
Procesamiento de datos

```
date/time: ,04/04/2019,07:09 p.m.
Run 1

time	V
0,000000	4,015237E-1
0,010000	4,090612E-1
0,020000	4,015237E-1
0,030000	3,990112E-1
```

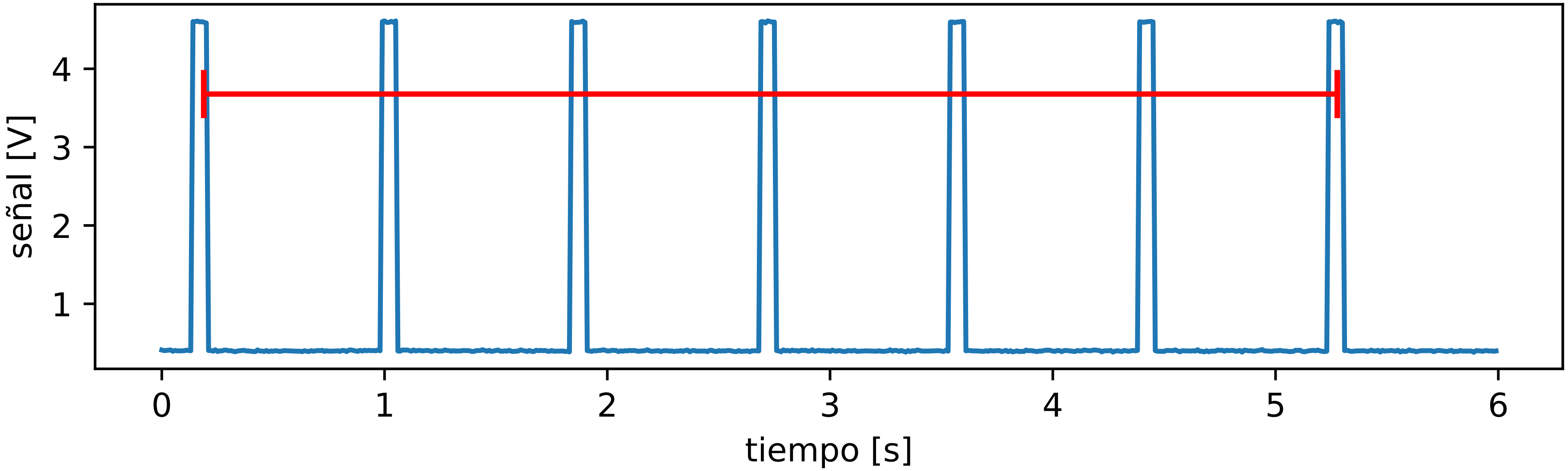
Improtar archivo de texto

¿Cómo procesamos los datos?

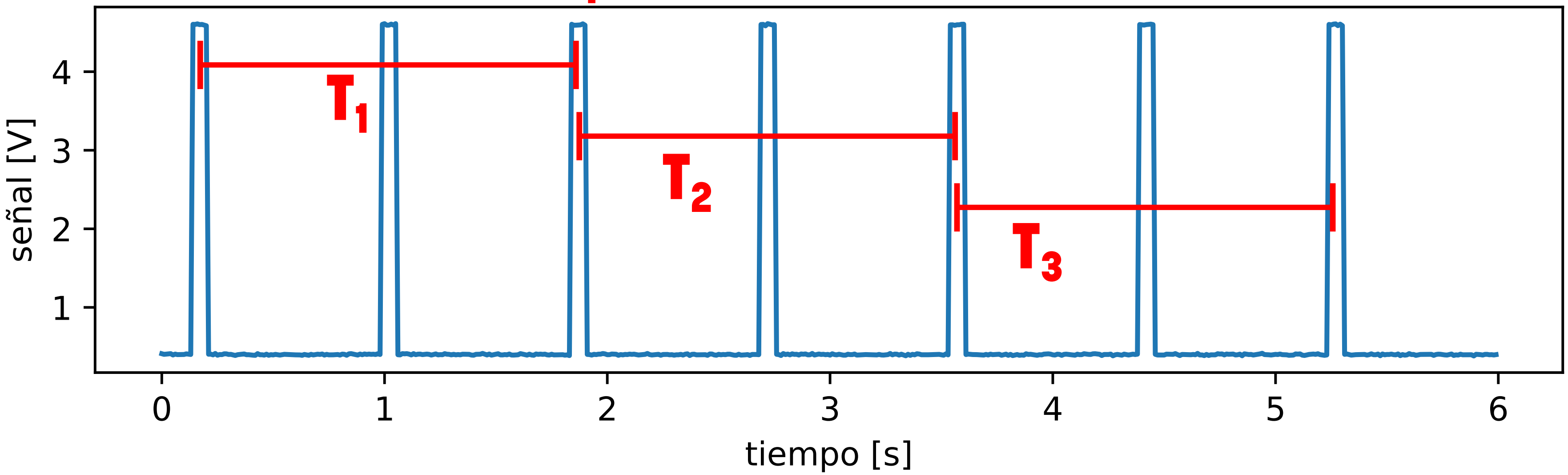


Diseño de la medición

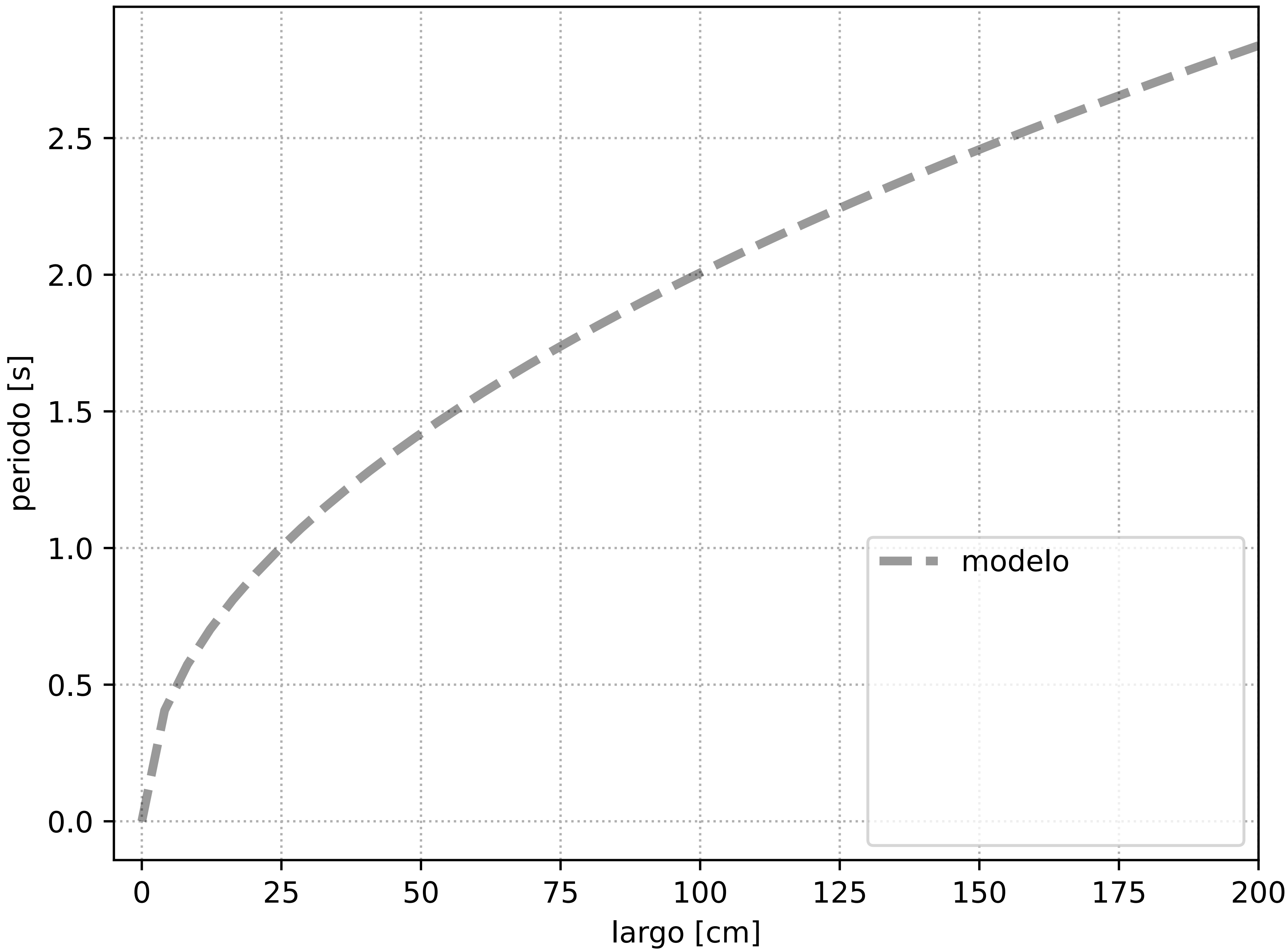
tiempo medido = N periodos → T=? , ΔT=?



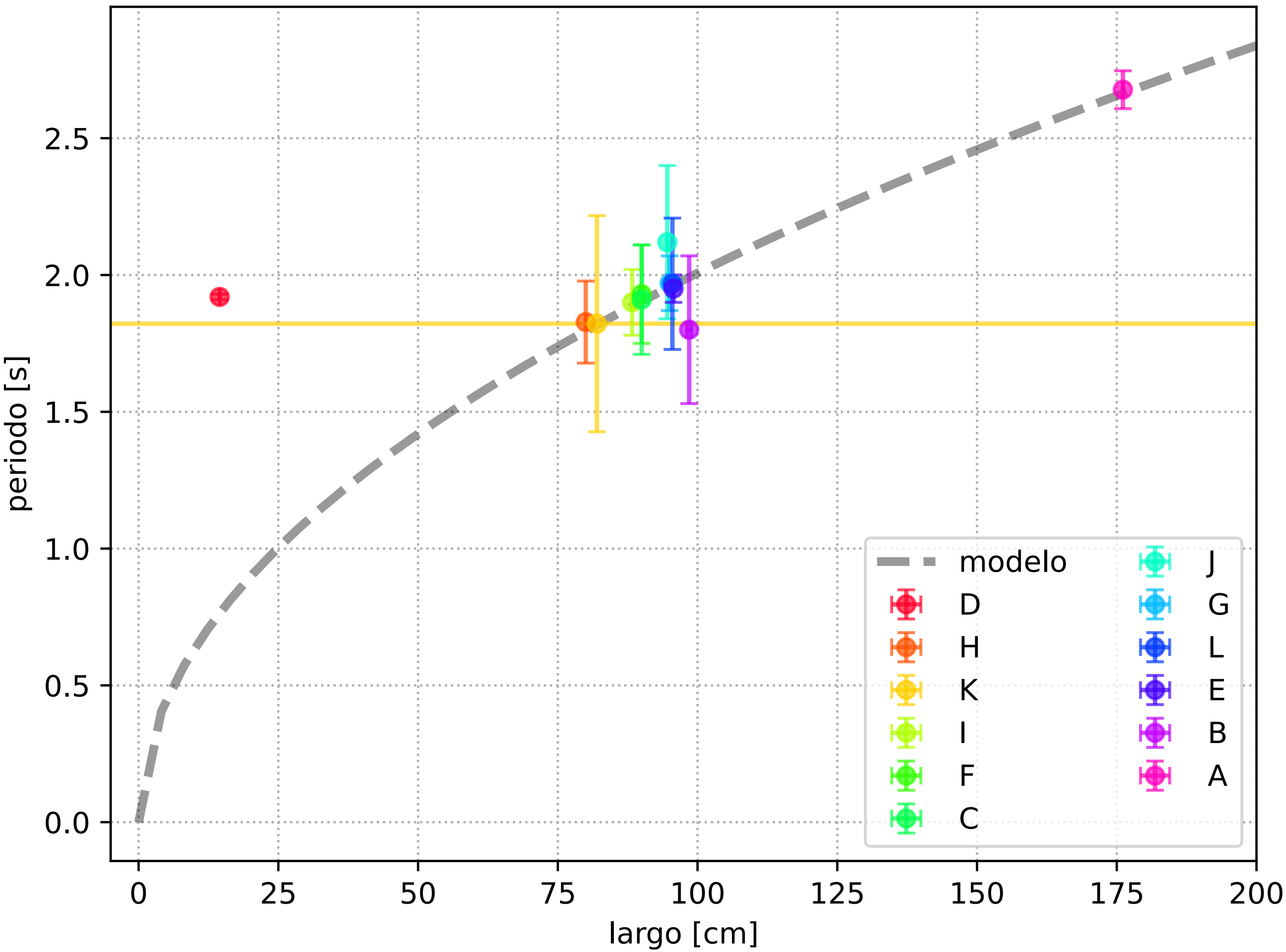
T=? , σ_T=?



Diseño experimento



Diseño experimento



Experimento

→ **Medir el periodo del péndulo para al menos 10 longitudes**
  → Planificar experimento: longitudes a medir, forma de medición
  → Medir con Photogate y procesar datos.
  → ¿Como se asigna el error del periodo?

→ **Graficar los datos, evaluar modelo y reportar medición de g**
  → Graficar con datos originales con sus incertezas
  → Linealizar los datos y realizar un ajuste mediante cuadrados mínimos
  → Analizar consistencia del modelo y bondad de ajuste
  → Reportar la medición indirecta de g con su incerteza