Clase 2

Comparación de datos con modelos (cuadrados mínimos)

https://asignaturas.df.uba.ar/myt-wisniacki/laboratorio/

Comparación de datos y modelos

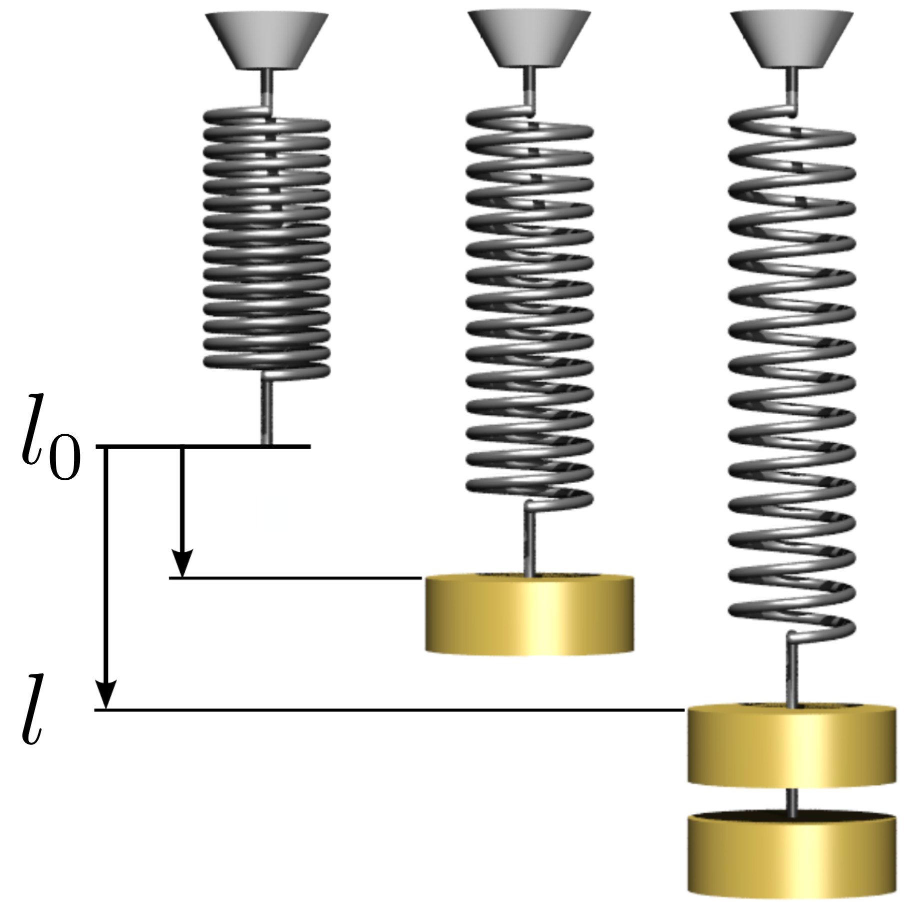
Ejemplo de modelo: Ley de Hook

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

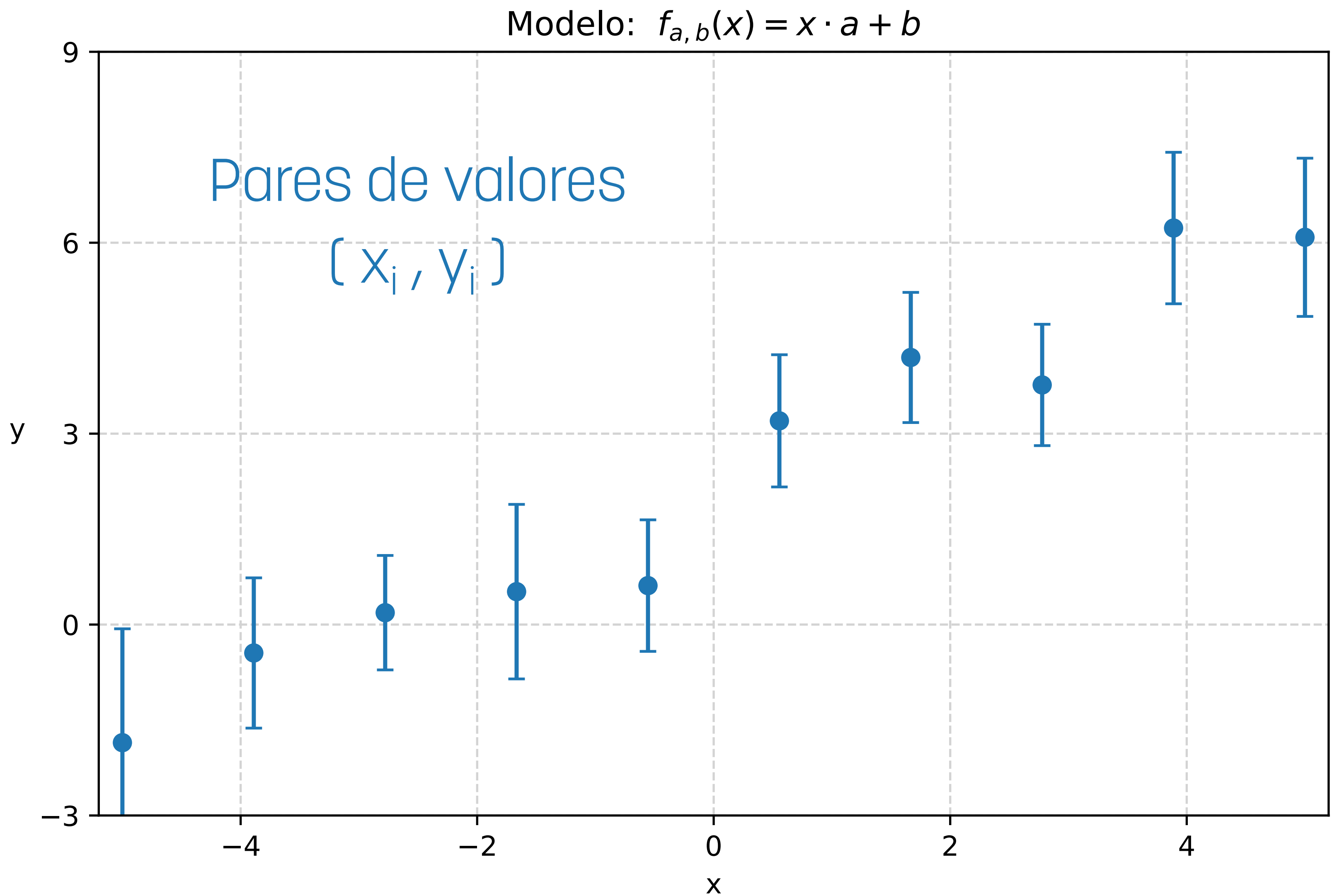
Conjunto de datos medidos (con sus errores)

Modelo Lineal

parámetros

f(x) = x a + b

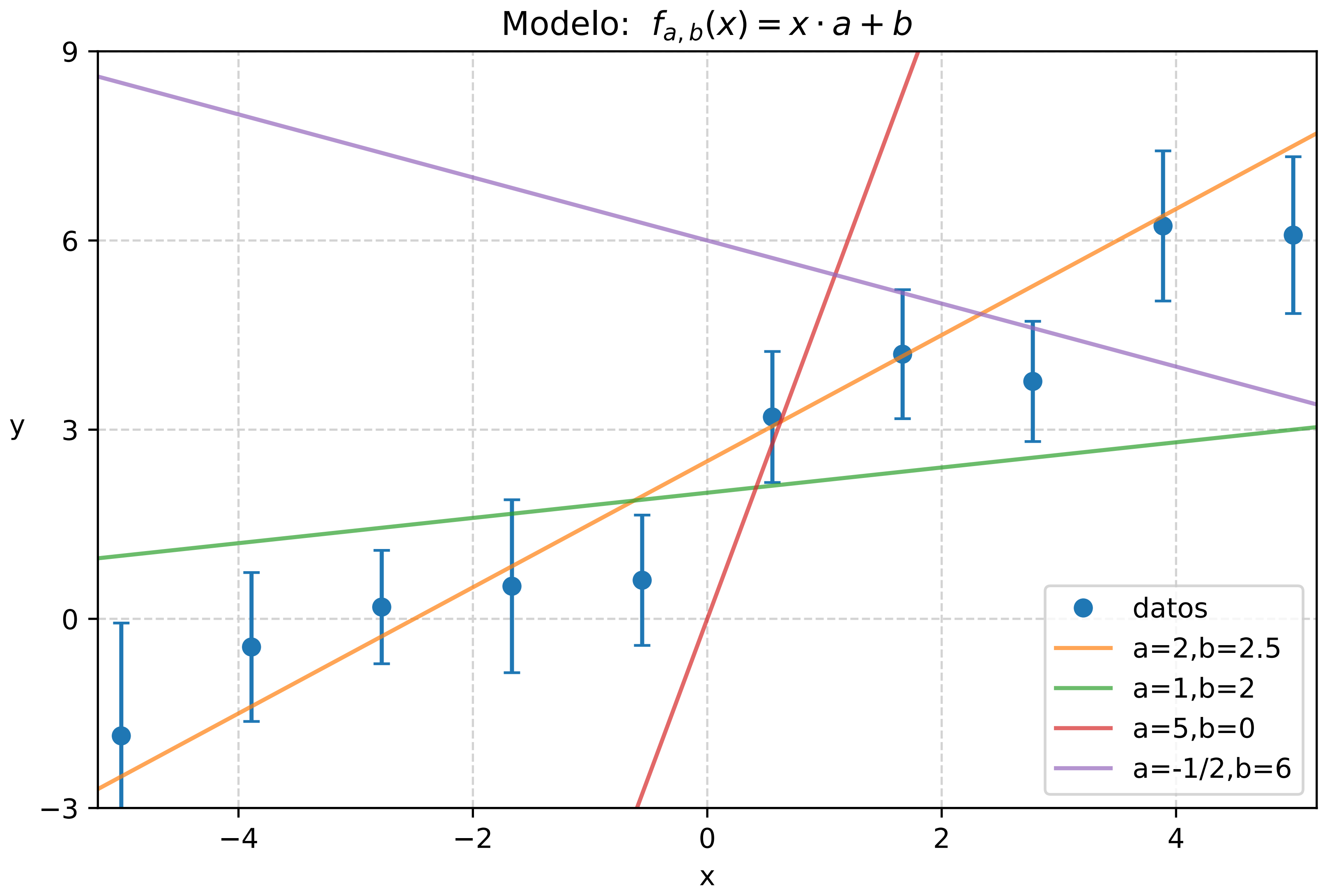
¿Responden los datos al modelo?



Cuadrados mínimos

Modelo para diferentes parámetros

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

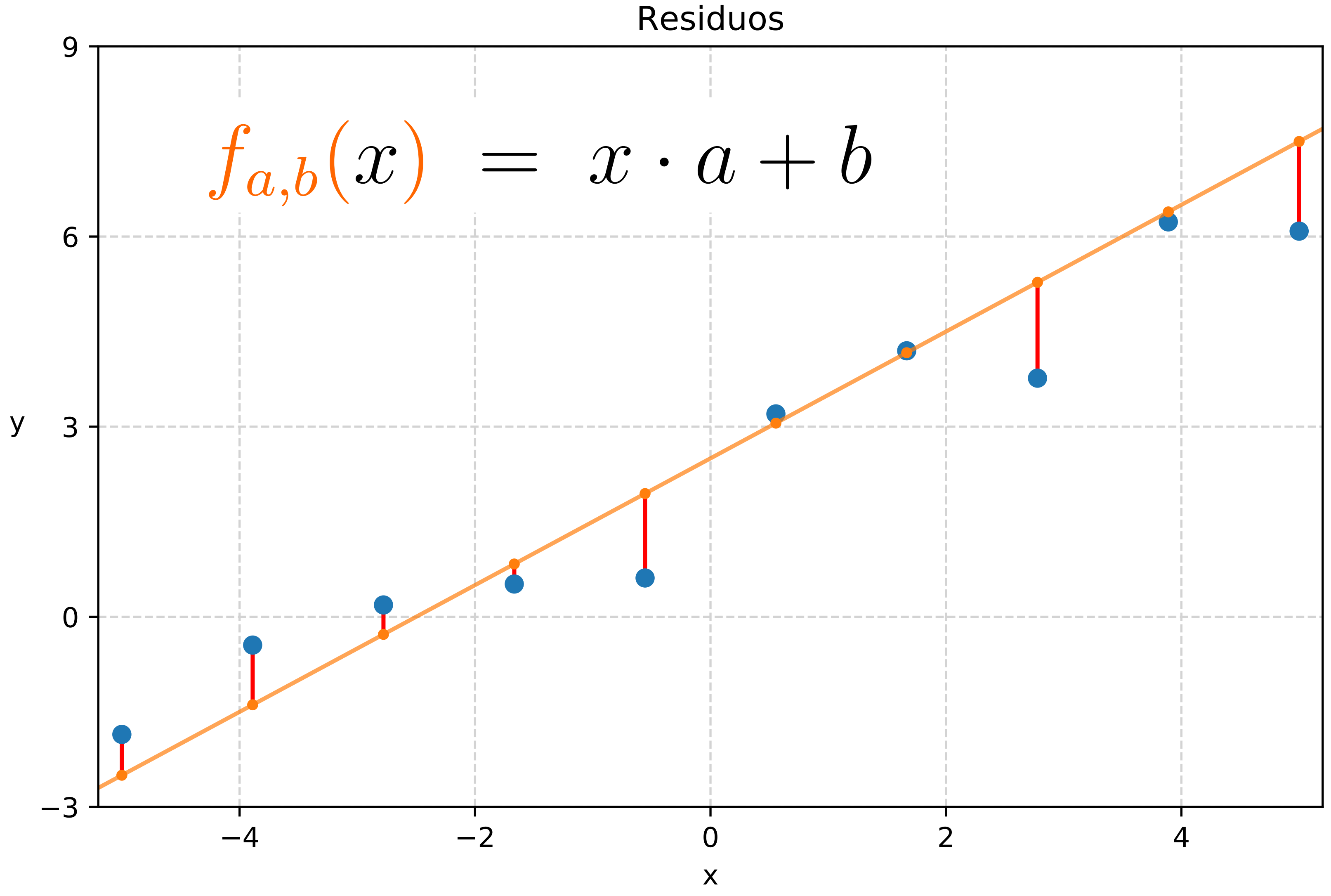


Cuadrados mínimos

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

$$(x_i, y_i)$$

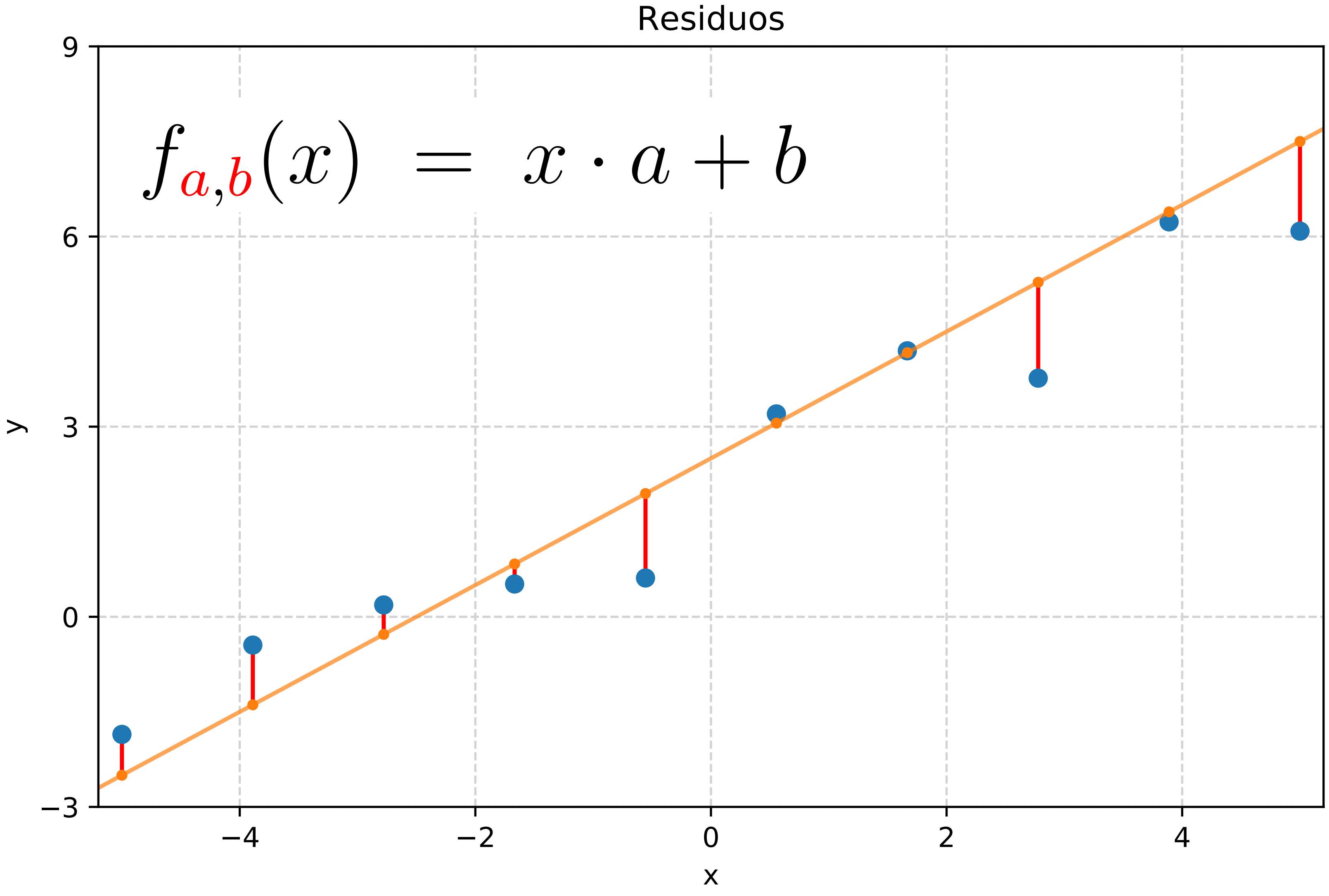
Fabricar una *"medida de distancia"* entre las predicciones del modelo y los datos medidos



Cuadrados mínimos

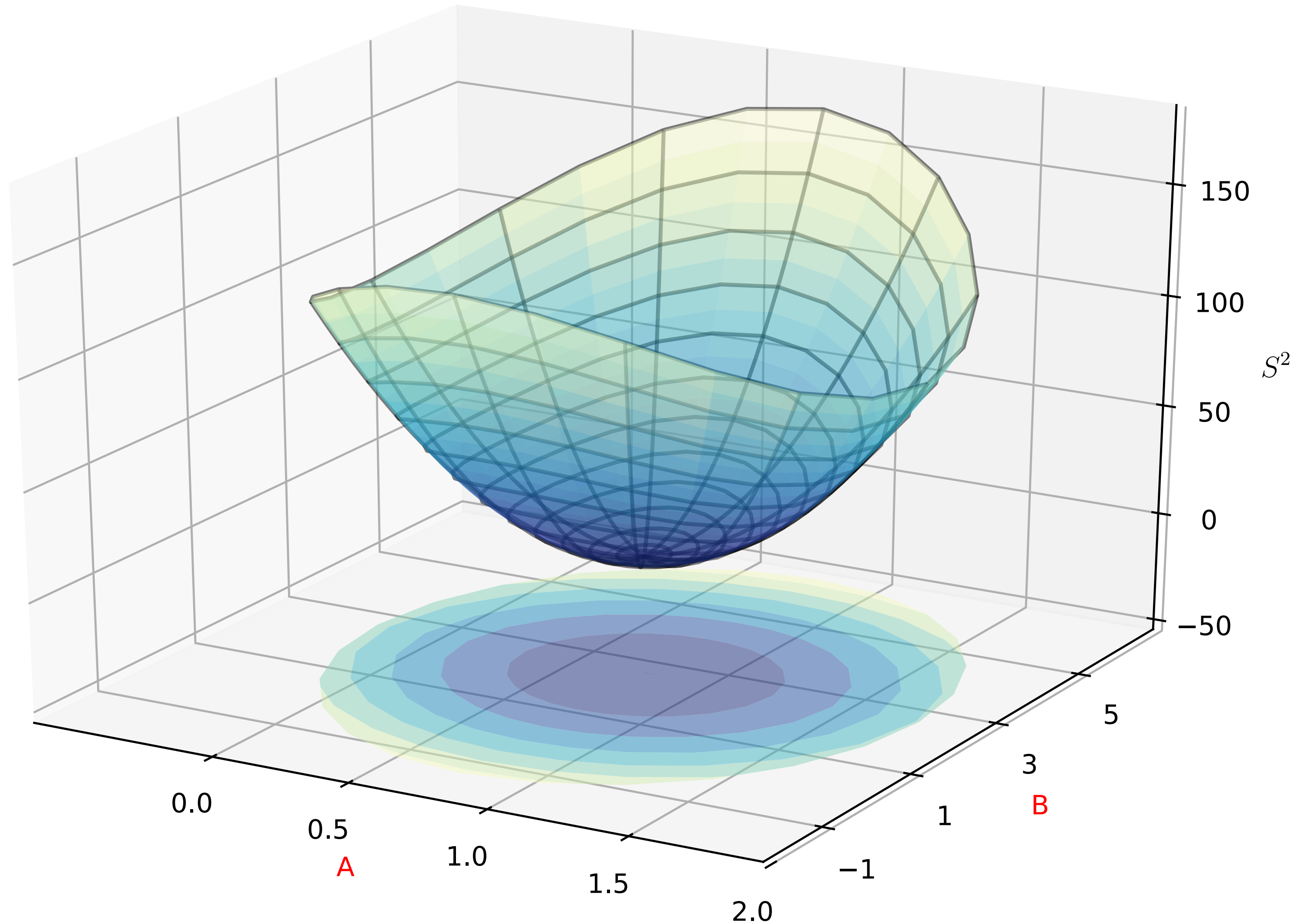
$$S^2(a,b) = \sum_{i=1}^{N} [y_i - f_{a,b}(x_i)]^2 = \sum_{i=1}^{N} [y_i - x_i \cdot a - b]^2$$

$$(x_i, y_i)$$



Cuadrados mínimos

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



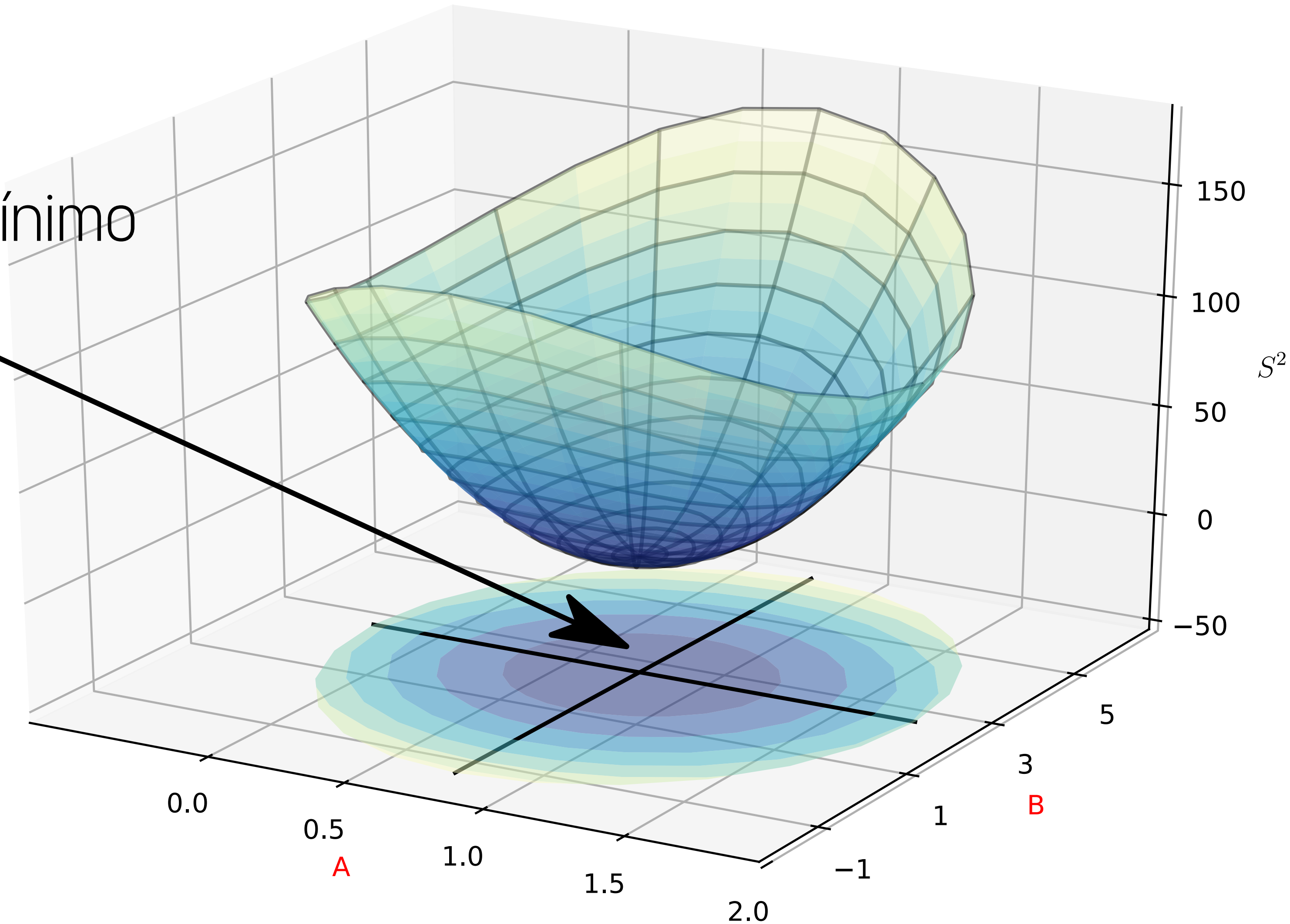
Superficie que representa a S^2

Cuadrados mínimos

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

Búsqueda del mínimo

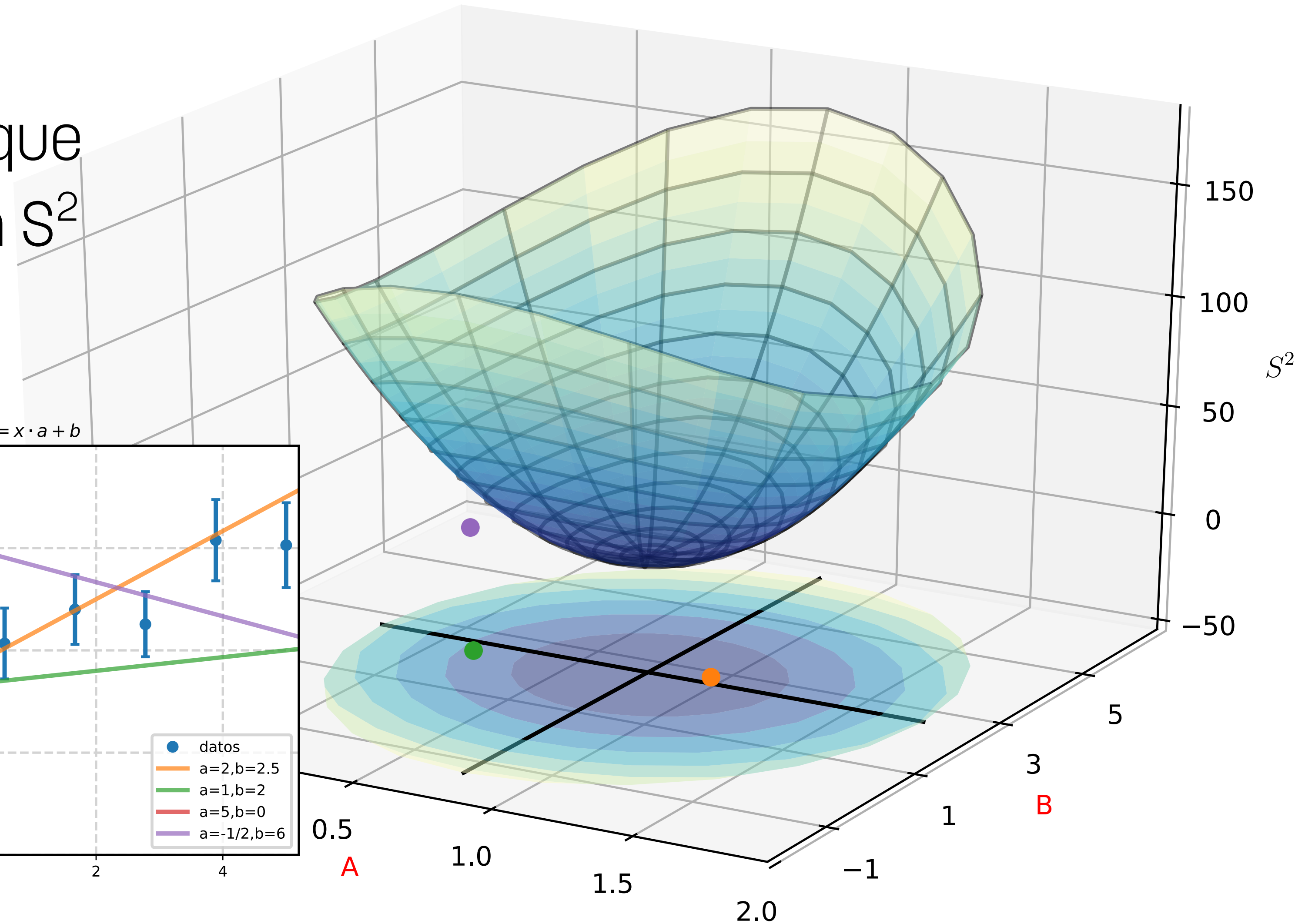
Algoritmo

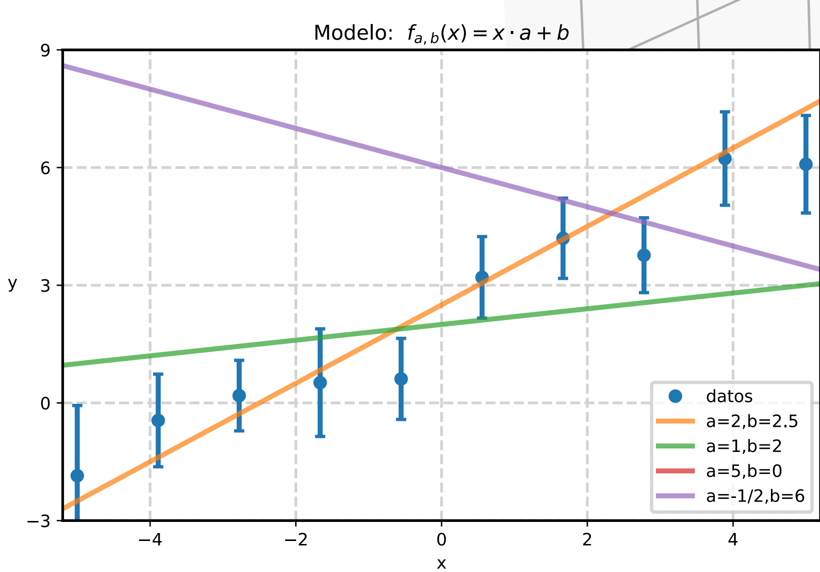


Cuadrados mínimos

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

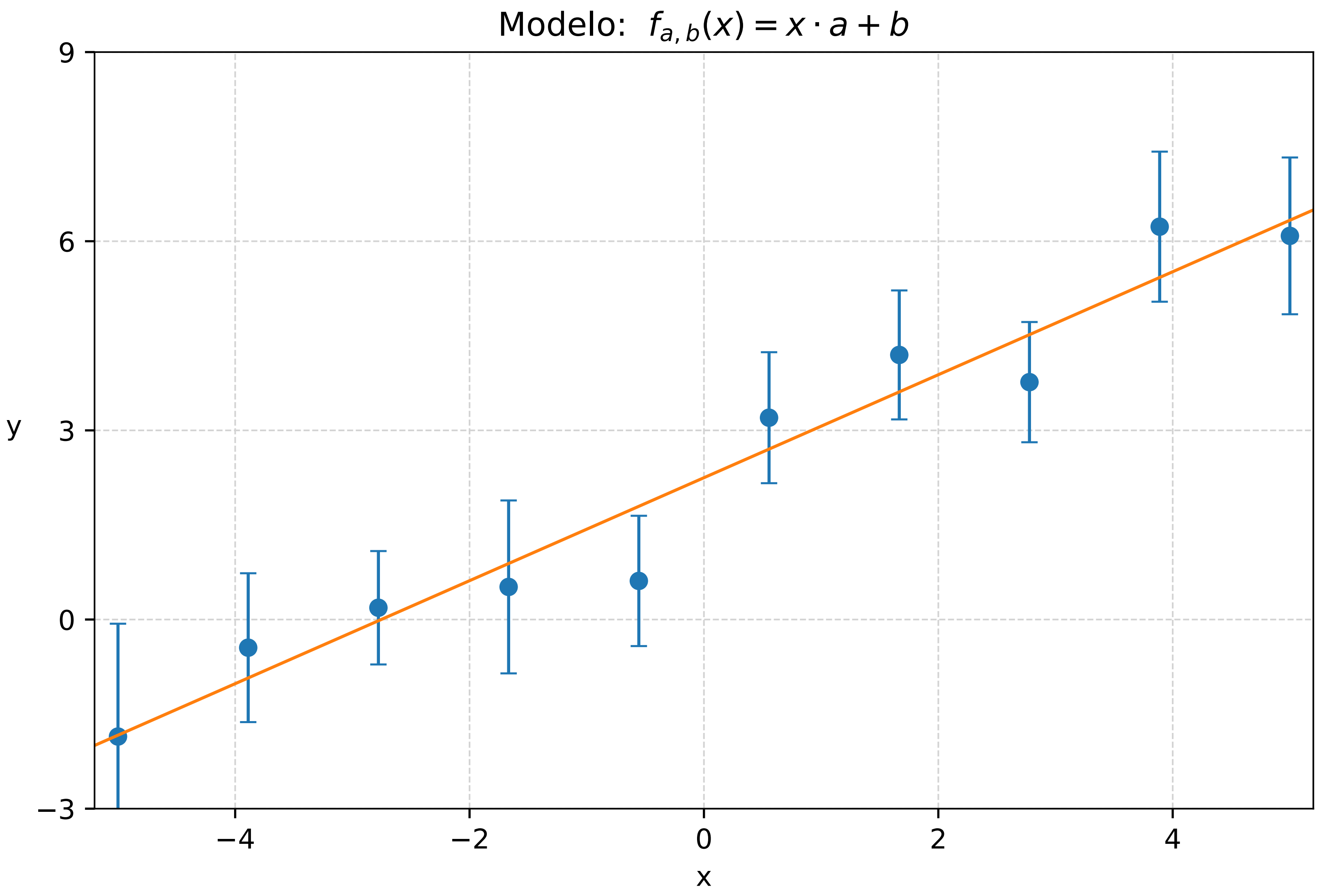
Ejemplos de parámetros que no minimizan S^2





Cuadrados mínimos

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



Cuadrados mínimos

Algoritmo para modelo lineal (sin incertezas en y)

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

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

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

Cuadrados mínimos

Algoritmo para modelo lineal (con incertezas en y)

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

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

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

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

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

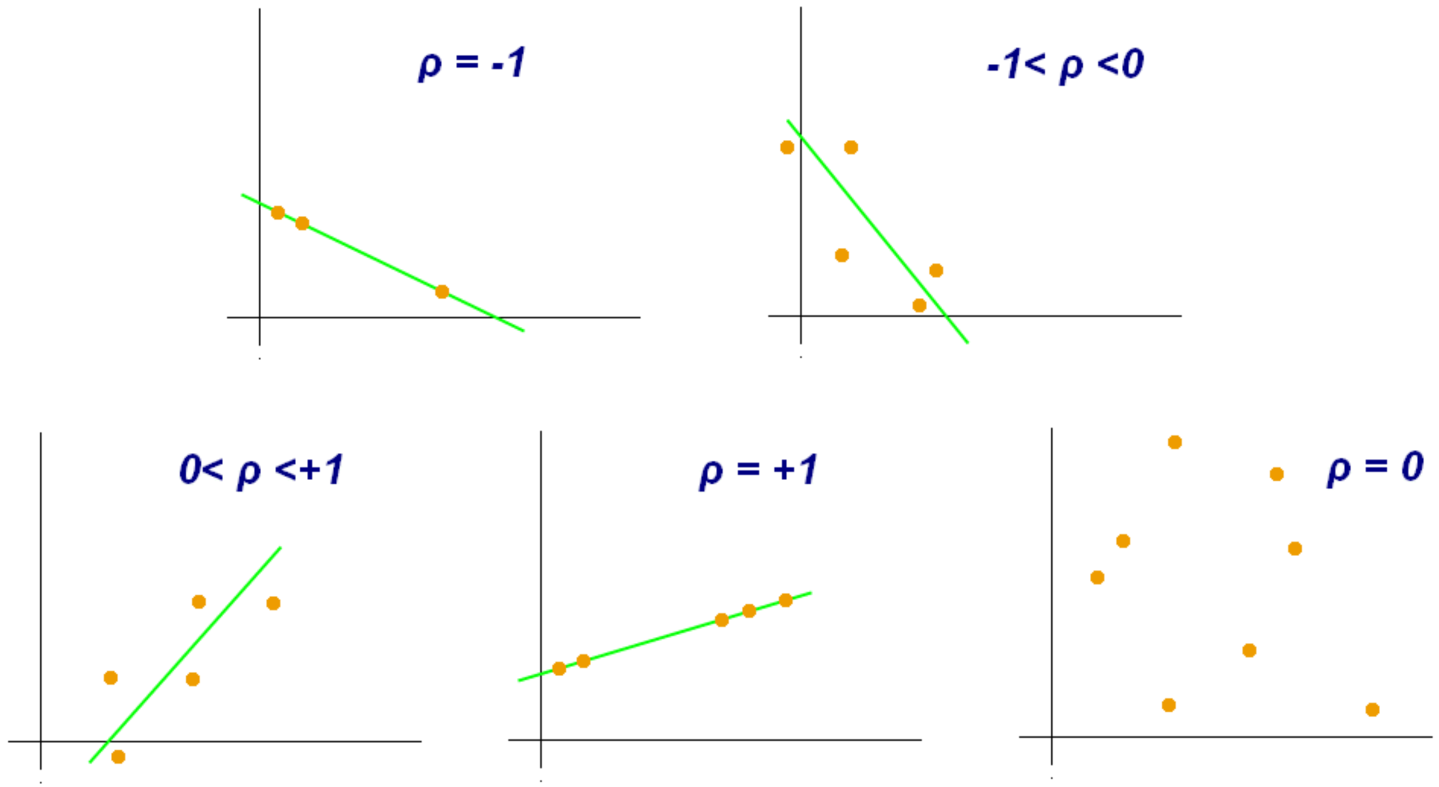
$$R^2 = 1 - \frac{\mathrm{RSS}}{\mathrm{SST}}$$

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

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

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

- **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

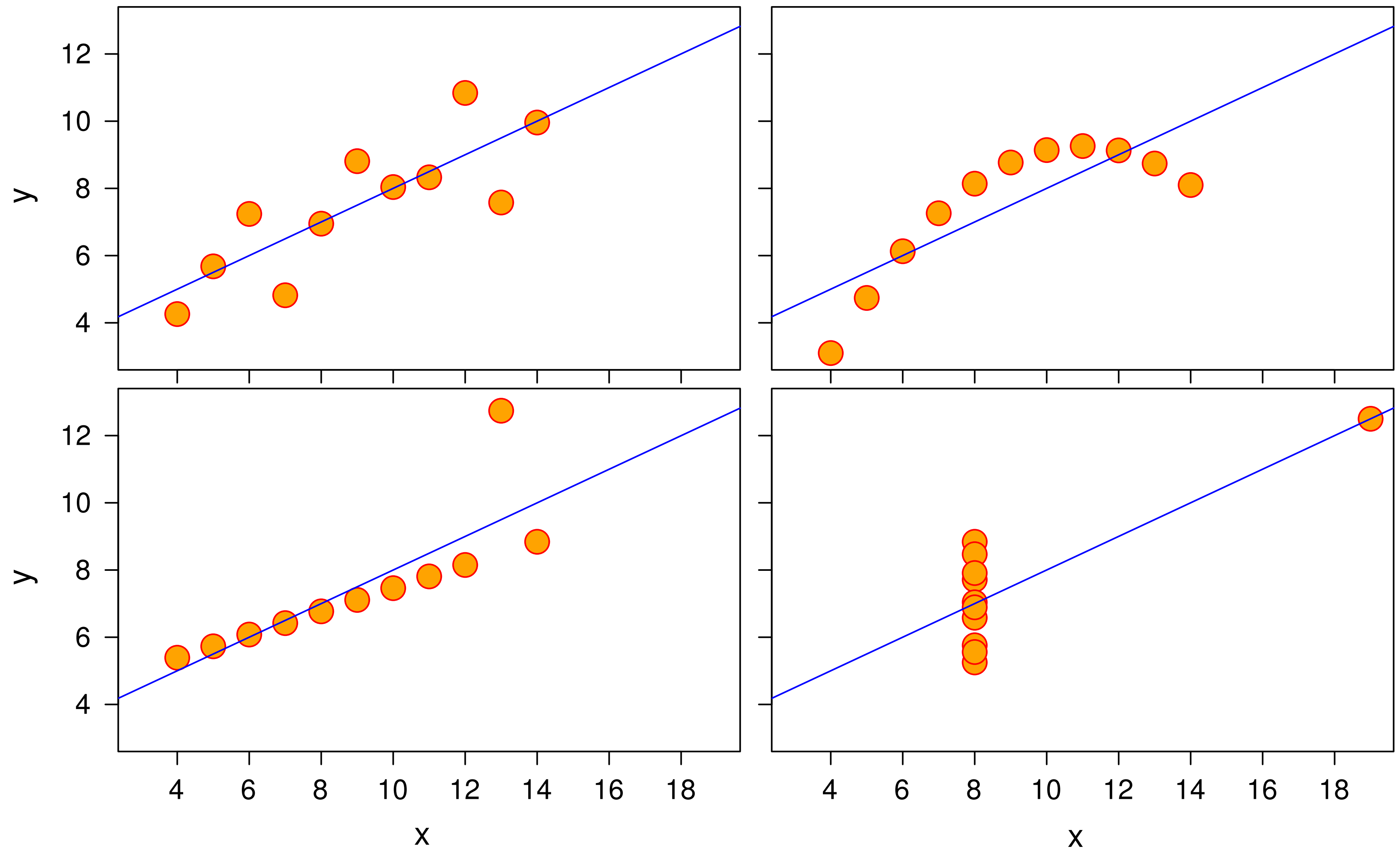


Cuarteto de Anscombe

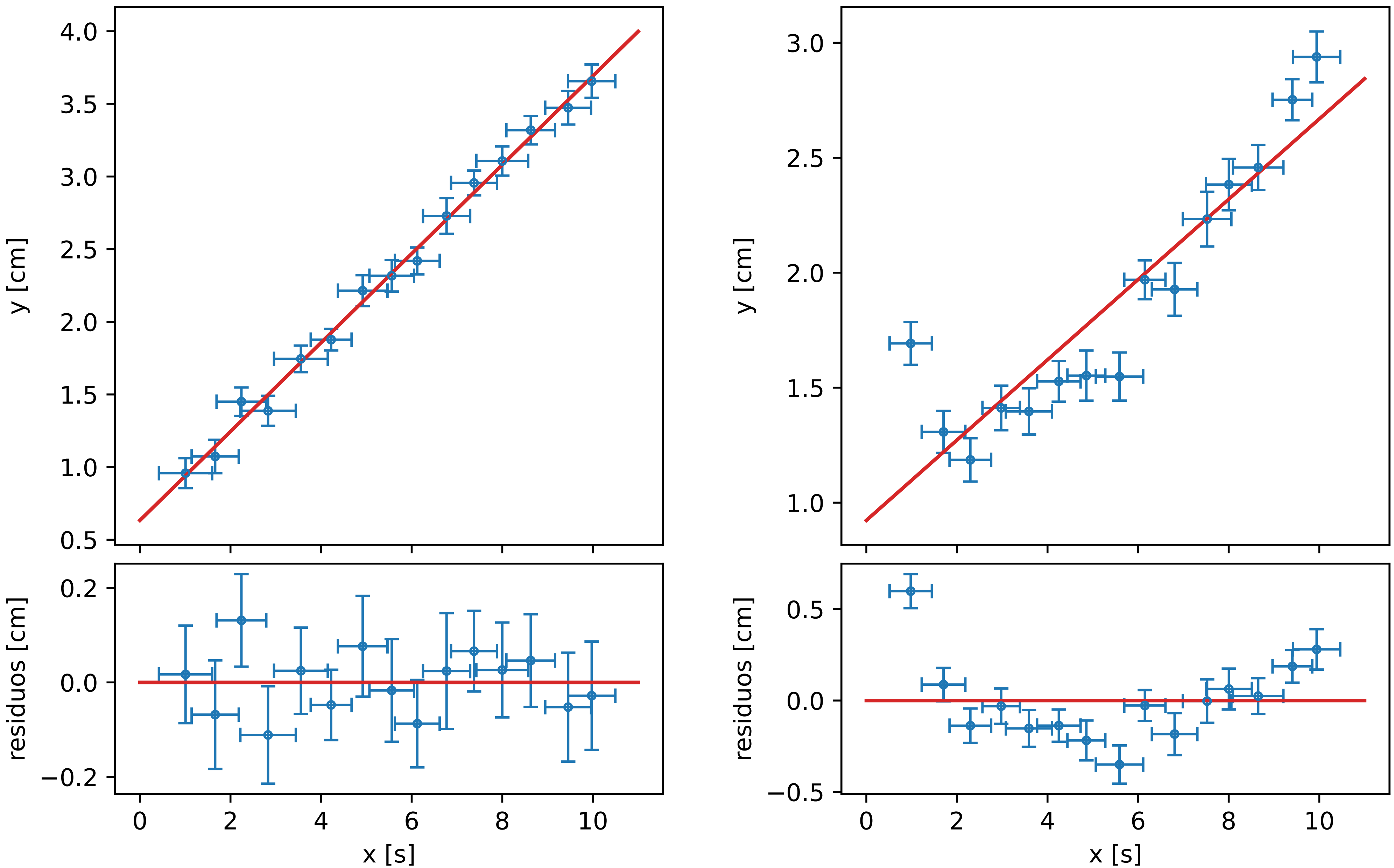
$$\bar{x} = 9$$
$$\mathrm{std}(x)^2 = 11$$
$$\bar{y} = 7.50$$
$$\mathrm{std}(y)^2 = 4.125$$
$$\mathrm{cor}(x,y) = 0.816$$
$$\mathrm{R}^2 = 0.67$$

linear fit

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



Residuos



¿...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)**

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

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

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

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

$$a \equiv 2$$

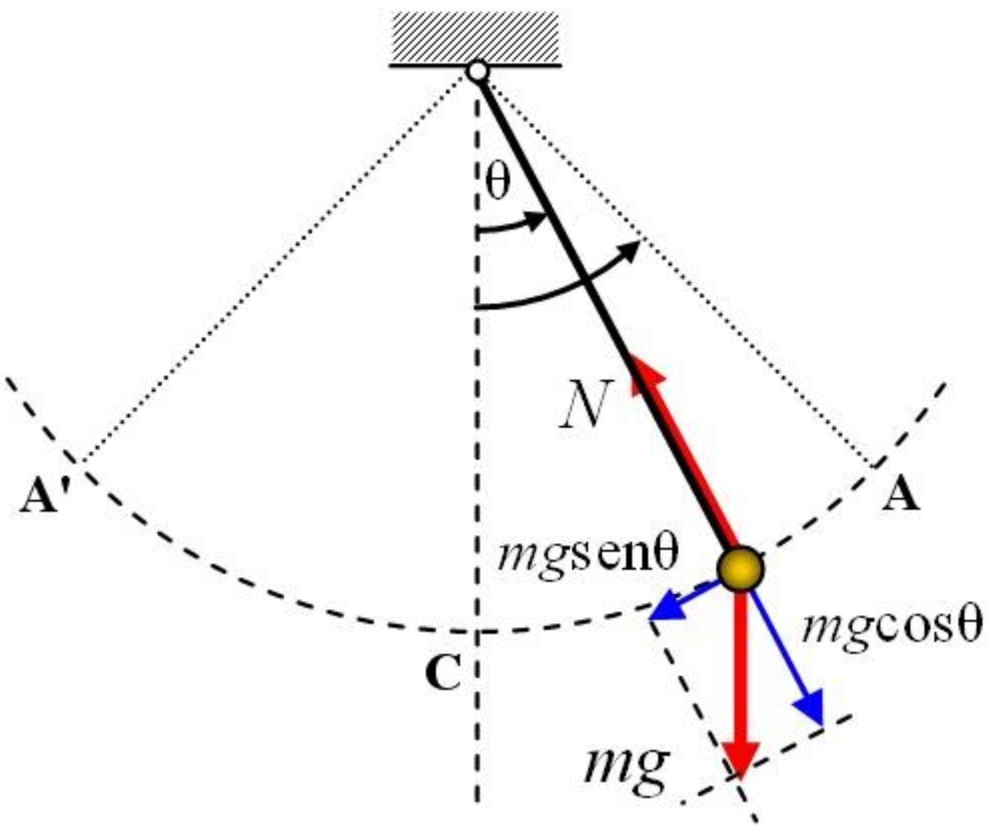
$$b \equiv \log\left(\frac{1}{2}g\right)$$

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

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

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:

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

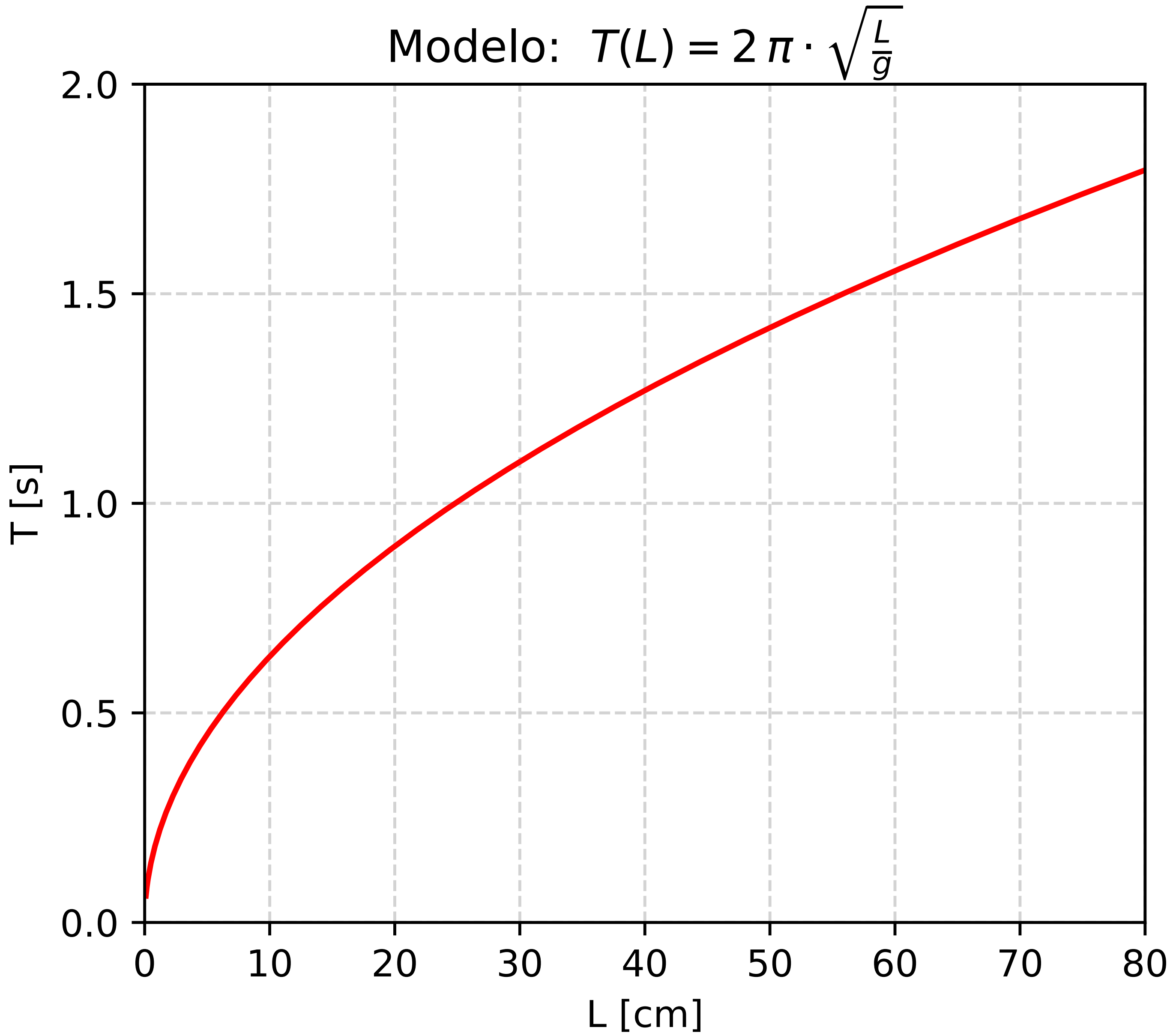
$$\omega = \frac{2\,\pi}{T} = \sqrt{\frac{g}{L}}$$

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

Actividad: validar modelo péndulo

Modelo de péndulo

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



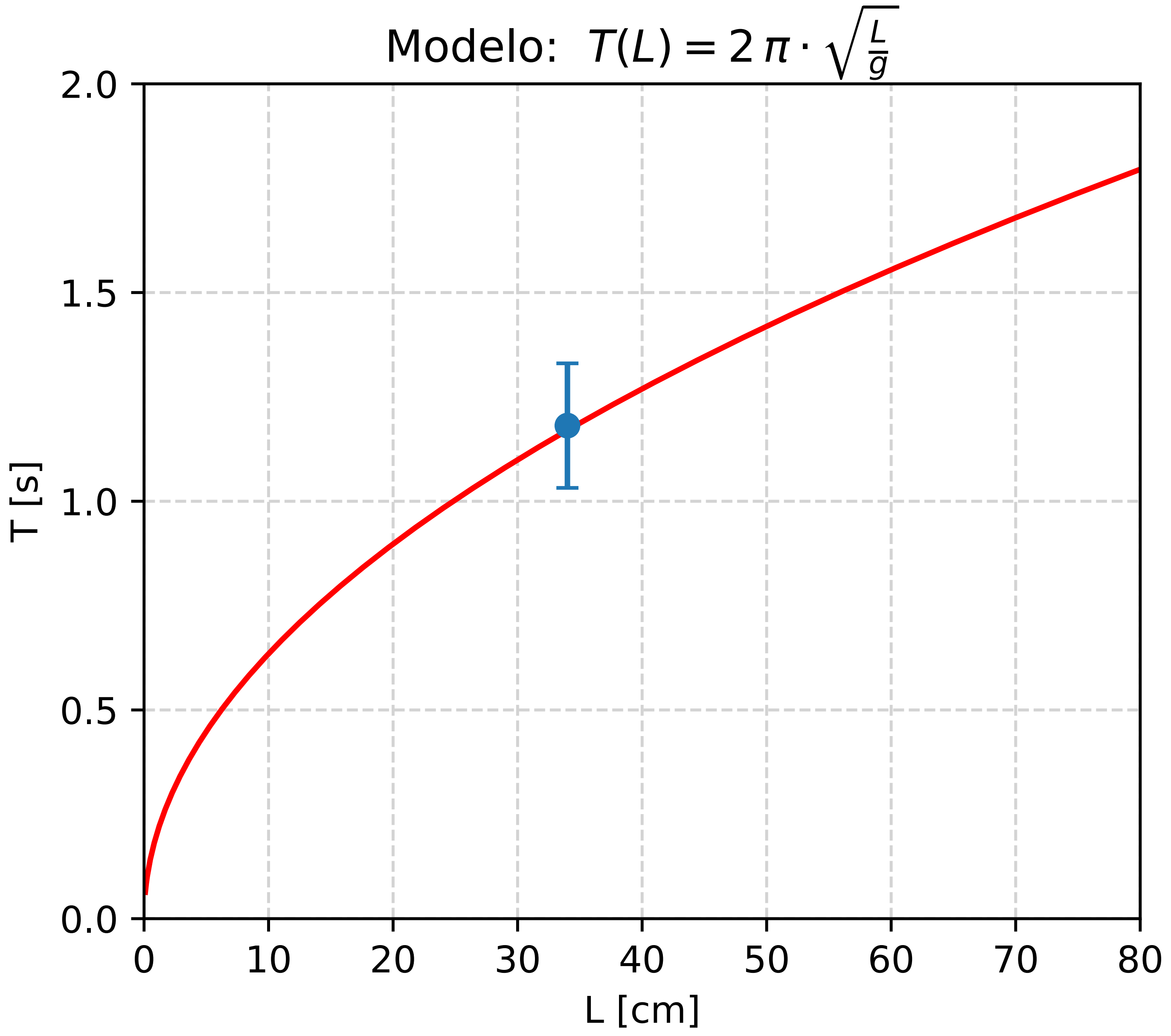
Actividad: validar modelo péndulo

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

$$(T_i, L_i)$$