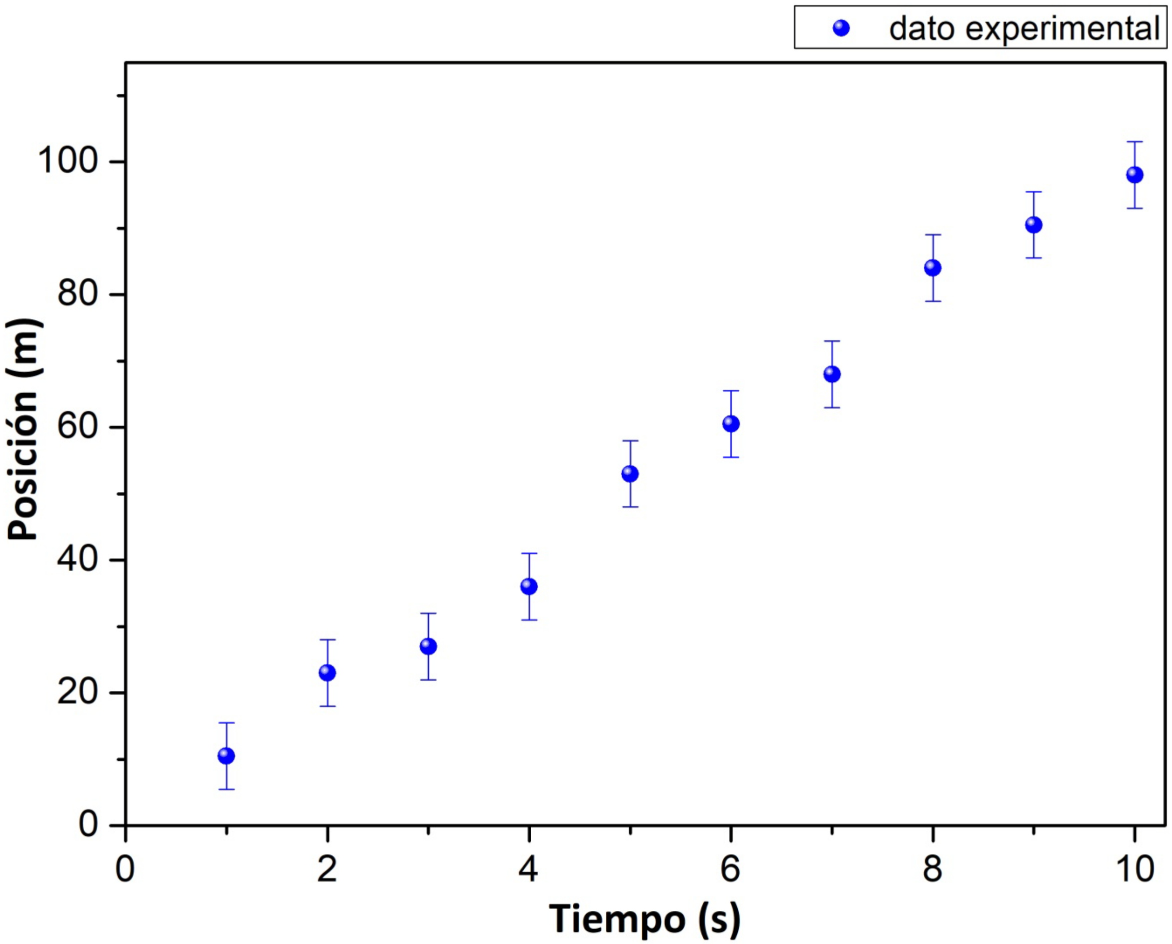
Guía 2: Estimación de la aceleración de la gravedad – Regresión lineal

Determinación de g a partir de la medición de del período de un péndulo de longitud variable

En muchos experimentos vamos a medir varios valores de dos variables físicas diferentes para investigar la relación matemática que hay entre ellas.

Ejemplo de estudio: Movimiento rectilíneo uniforme (MRU)



Mediciones:

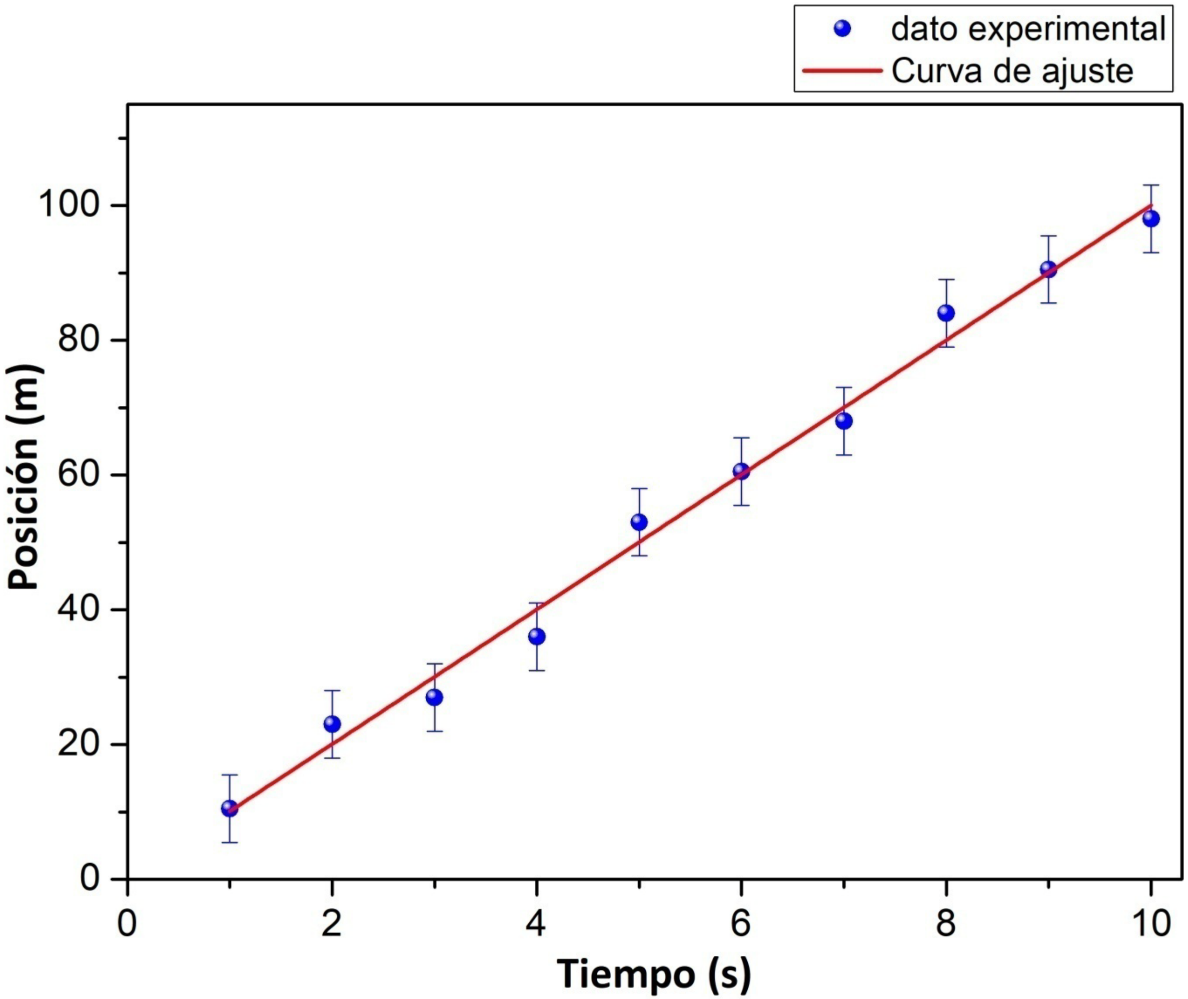
$$(x_i \pm \Delta x_i, y_i \pm \Delta y_i)$$

$x \rightarrow$ tiempo
$y \rightarrow$ posición

Los datos parecen seguir una relación lineal.

Modelo lineal:

$$f(x) = A.x + B$$



Mediciones:

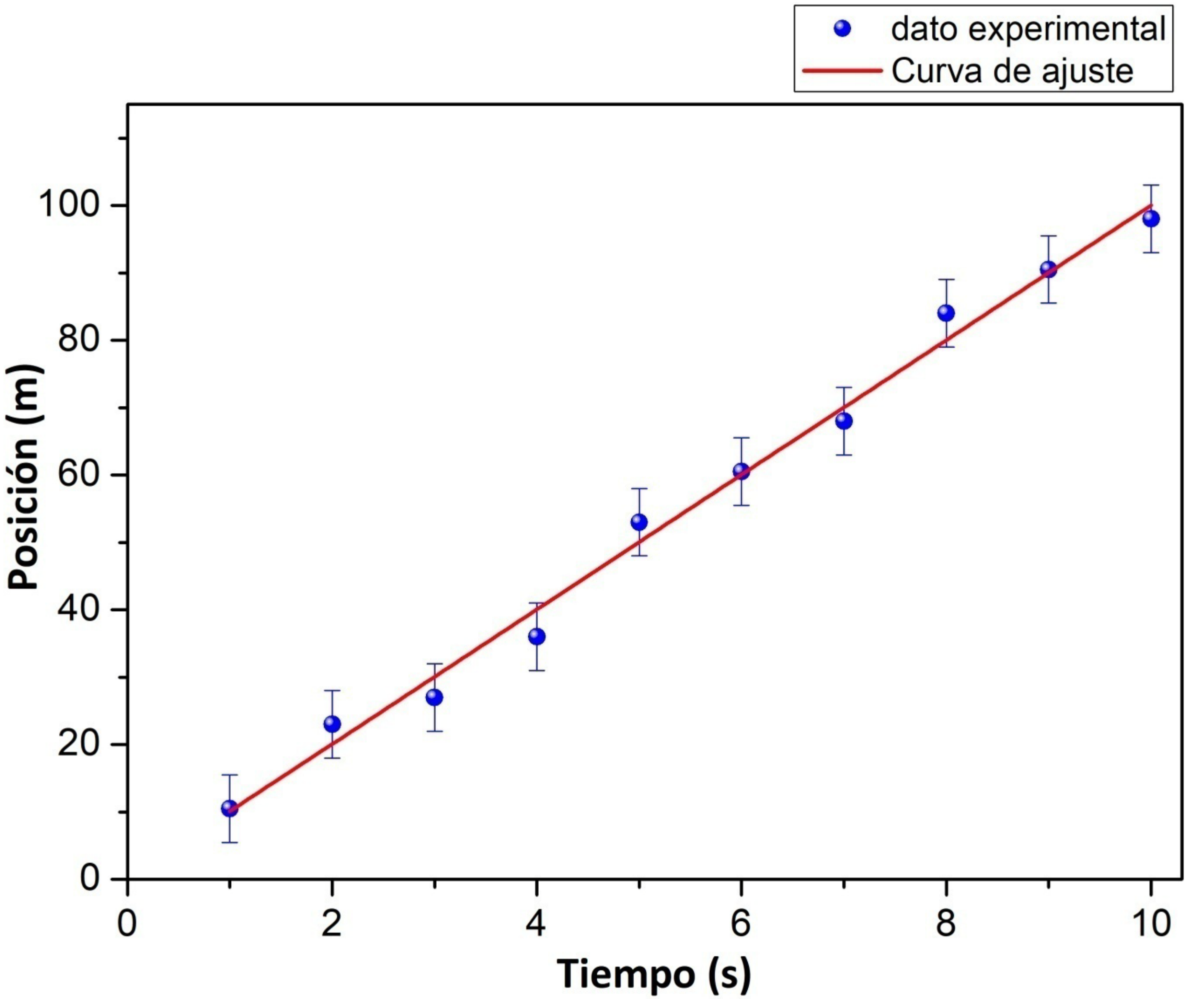
$$(x_i \pm \Delta x_i, y_i \pm \Delta y_i)$$

$x \rightarrow$ tiempo
$y \rightarrow$ posición

Modelo lineal:

$$f(x) = A.x + B$$

¿Responden los datos al modelo?



Mediciones:

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

$x \rightarrow$ tiempo
$y \rightarrow$ posición

Modelo lineal:

$$f(x) = A.x + B$$

¿Responden los datos al modelo?

Sabemos que en MRU, la posición de un objeto se describe por la ecuación:

$$d = d_0 + v_0.t$$

¿Cuál es la recta que mejor representa la relación? ¿Valores de A y B?

¿Cómo la obtenemos?

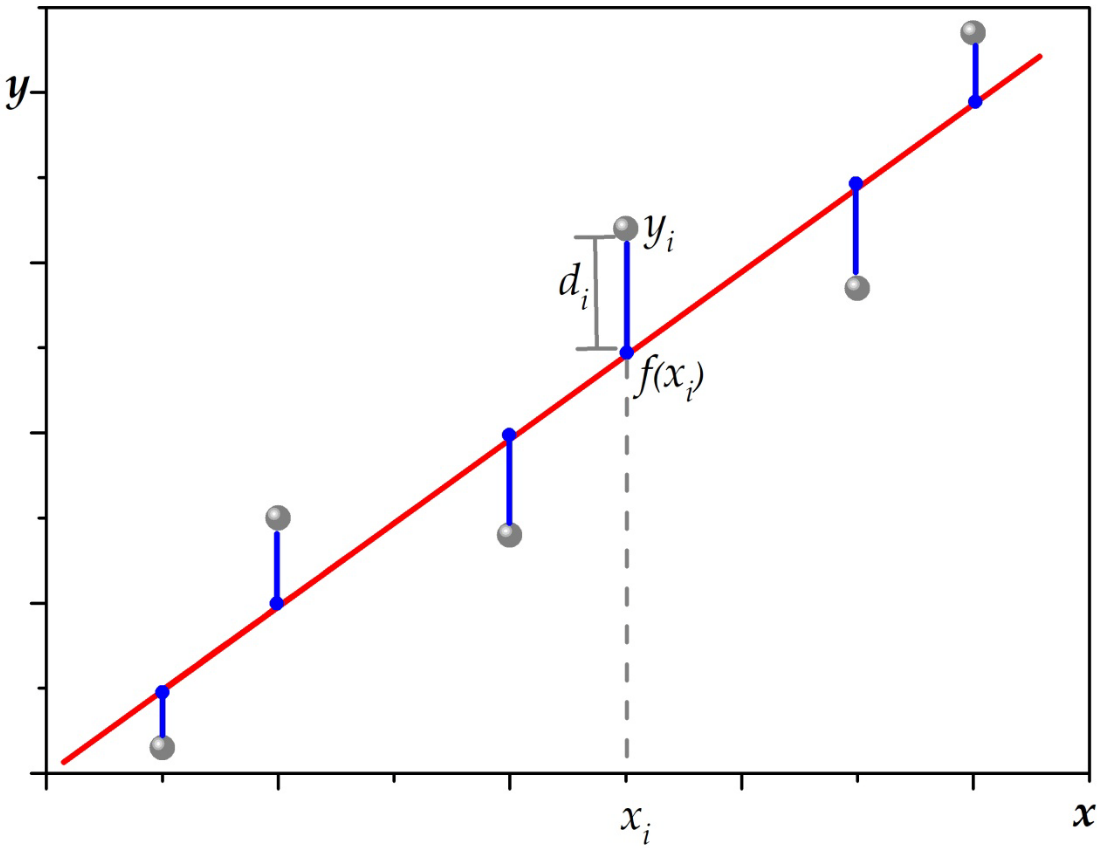
¿Cómo la informamos?

Uno de los métodos más usados para determinar la mejor recta que pasa entre varios puntos experimentales es el **método de regresión lineal por cuadrados mínimos**.

Regresión lineal por cuadrados mínimos <u>ordinarios</u> $\rightarrow$ **se supone despreciables las incertezas Δx_i y Δy_i.**

Uno de los métodos más usados para determinar la mejor recta que pasa entre varios puntos experimentales es el **método de regresión lineal por cuadrados mínimos**.

Regresión lineal por cuadrados mínimos <u>ordinarios</u> → se supone despreciables las incertezas Δx_i y Δy_i.



Modelo lineal:

$$f(x) = A.x + B$$

y_i → valor medido

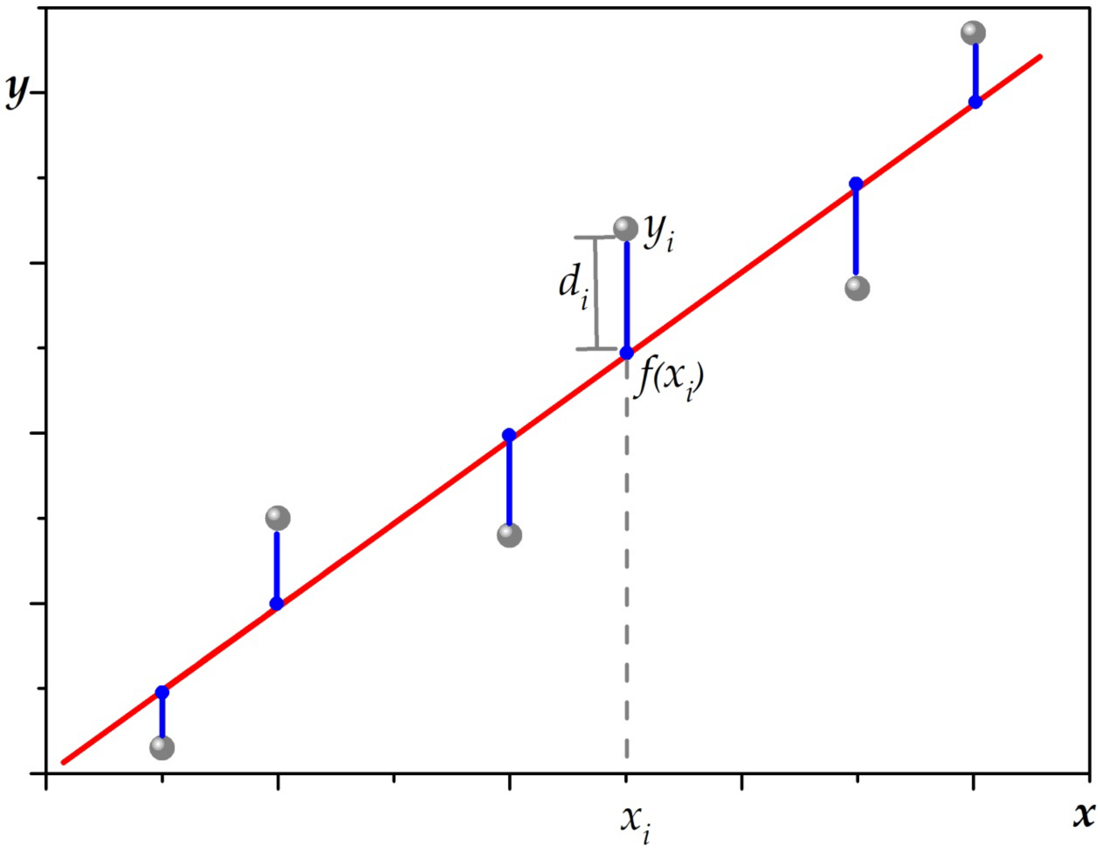
→ recta de regresión estimada

Residuo → diferencia entre el valor observado y los valores del modelo:

$$d_i = y_i - f(x_i)$$
$$d_i = y_i - A.x_i - B$$

Uno de los métodos más usados para determinar la mejor recta que pasa entre varios puntos experimentales es el **método de regresión lineal por cuadrados mínimos**.

Regresión lineal por cuadrados mínimos ordinarios → se supone despreciables las incertezas Δx_i y Δy_i.



Modelo lineal:

$$f(x) = A.x + B$$

y_i → valor medido

→ recta de regresión estimada

Residuo → diferencia entre el valor observado y los valores del modelo:

$$d_i = y_i - f(x_i)$$
$$d_i = y_i - A.x_i - B$$

El método de cuadrados mínimos busca los parámetros A y B que minimizan las distancias verticales d_i entre los puntos medidos y los valores de recta de regresión estimada.

Si tengo N mediciones $(x_i \pm \Delta x_i, y_i \pm \Delta y_i)$ $\rightarrow$ Cuadrados mínimos **ordinarios**

$$M = \sum_{i=1}^{N} (d_i)^2 = \sum_{i=1}^{N} (y_i - A.x_i - B)^2$$

$\rightarrow$ M sea un mínimo

(minimizamos la suma de los residuos d_i al cuadrado)

$$\frac{\partial M}{\partial A} = 0 \qquad \frac{\partial M}{\partial B} = 0$$

Si tengo N mediciones $(x_i \pm \Delta x_i, y_i \pm \Delta y_i)$ → Cuadrados mínimos **ordinarios**

$$M = \sum_{i=1}^{N} (d_i)^2 = \sum_{i=1}^{N} (y_i - A.x_i - B)^2$$

→ M sea un mínimo

(minimizamos la suma de los residuos d_i al cuadrado)

$$\frac{\partial M}{\partial A} = 0 \qquad \frac{\partial M}{\partial B} = 0$$

Se obtienen los parámetros del modelo:

$$A = \frac{N \sum (x_i\, y_i) - (\sum x_i) \cdot (\sum y_i)}{N \sum x_i^2 - (\sum x_i)^2} \qquad B = \frac{(\sum x_i^2)\,(\sum y_i) - (\sum x_i) \cdot (\sum x_i\, y_i)}{N \sum x_i^2 - (\sum x_i)^2}$$

Para regresiones lineales, los valores de A y B (**y sus errores**) se determinan de forma **unívoca**.

Incertezas →

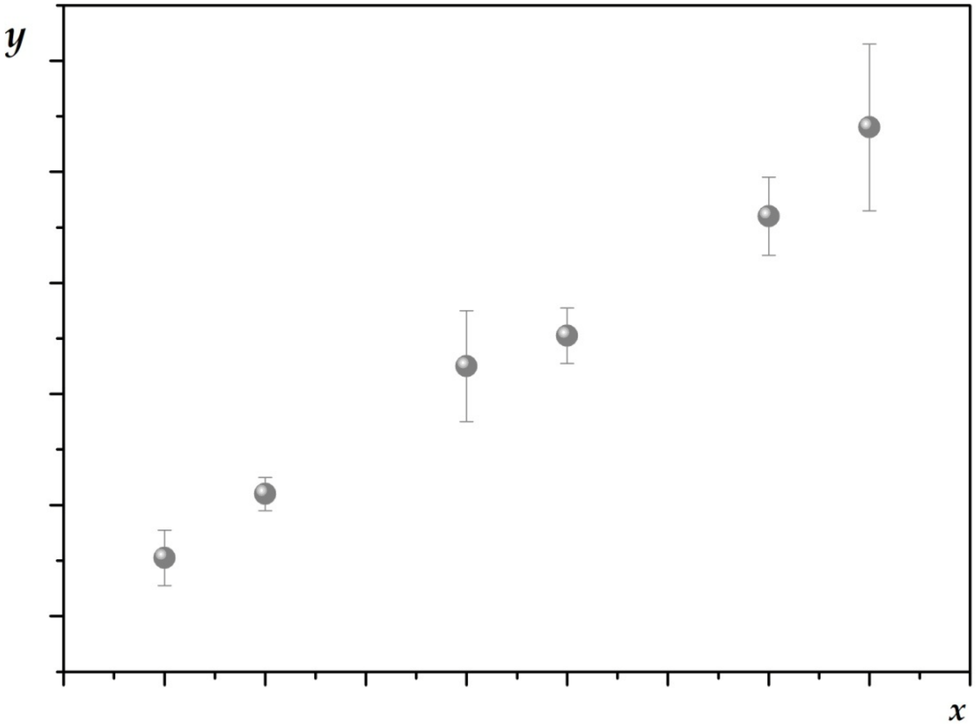
$$\sigma_A = \sqrt{\frac{N \sum (y_i - A.x_i - B)^2}{(N-2)\left[N \sum x_i^2 - (\sum x_i)^2\right]}}$$

$$\sigma_B = \sqrt{\left[1 + \frac{(\sum x_i)^2}{N \sum x_i^2 - (\sum x_i)^2}\right] \frac{\sum (y_i - A.x_i - B)^2}{N(N-2)}}$$

Detalles: D.C. Baird, *Experimentación*,

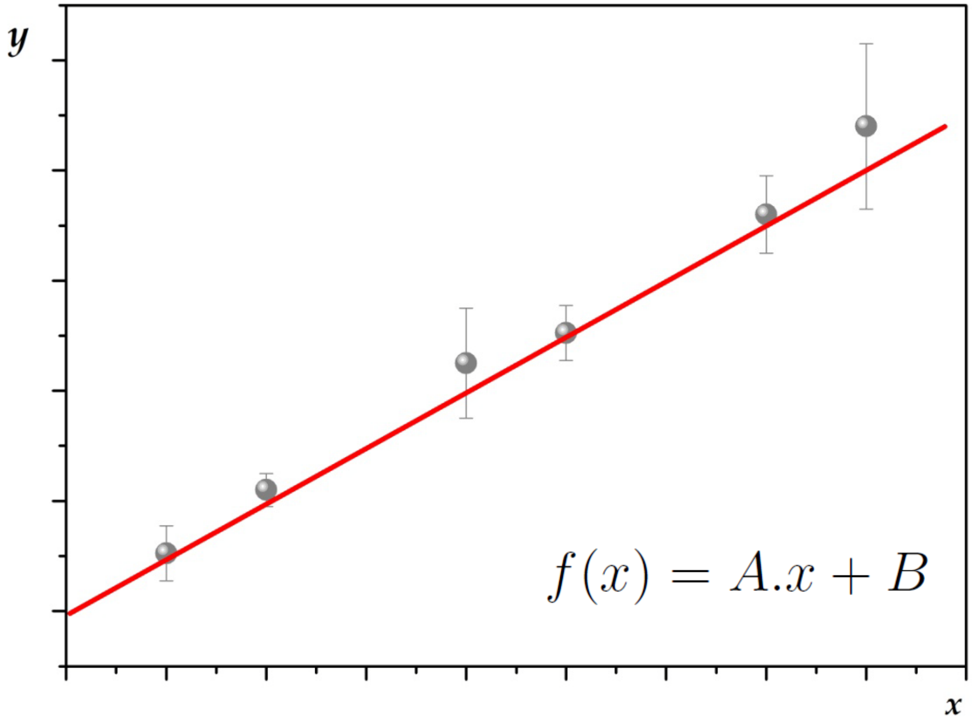
¿Todos los puntos son equivalentes? ¿Confiamos más en alguno que en otro?

Regresión lineal por cuadrados mínimos ponderados



¿Todos los puntos son equivalentes? ¿Confiamos más en alguno que en otro?

Regresión lineal por cuadrados mínimos ponderados



N mediciones:

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

Suposición: errores Δx_i despreciables frente a los errores Δy_i (variable x medida con mayor precisión que la variable y).

Idea general: procedimiento de minimización $\rightarrow$ se asigna mayor importancia a los d_i provenientes de valores de los y_i que tengan errores más chicos.

$\rightarrow$

Ponderar $\rightarrow$ multiplicar a cada d_i por un factor de peso (una cantidad que sea mayor cuanto más pequeño sea el correspondiente Δy_i).

$$\sum_{i=1}^{N}\left(\frac{d_i}{\Delta y_i}\right)^2=\sum_{i=1}^{N}\left(\frac{y_i-A.x_i-B}{\Delta y_i}\right)^2$$ → minimizar → Se obtienen expresiones para los parámetros ***A*** y ***B***, y sus incertezas **Δ*A*** y **Δ*B*** del modelo.

Definimos las cantidades w_i y Δ:

$$w_i=\frac{1}{\Delta y_i^2}\qquad\qquad \Delta=(\sum w_i).(\sum w_i.x_i^2)-(\sum w_i.x_i)^2$$

$$A=\frac{(\sum w_i).\sum(w_i.x_i.y_i)-(\sum w_i.x_i).(\sum w_i.y_i)}{\Delta}$$

$$B=\frac{(\sum w_i.x_i^2).(\sum w_i.y_i)-(\sum w_i.x_i).(\sum w_i.x_i.y_i)}{\Delta}$$

Regresión lineal por cuadrados mínimos ponderados.

Incertezas:

$$\sigma_A=\sqrt{\frac{\sum w_i}{\Delta}}\qquad\qquad \sigma_B=\sqrt{\frac{\sum(w_i.x_i^2)}{\Delta}}$$

Pendiente → $A\pm\sigma_A$

Ordenada al origen → $B\pm\sigma_B$

Coeficiente de determinación R^2 (R-Squared)

$$\boxed{R^2 = \frac{\left[\sum(x_i-\bar{x})\sum(y_i-\bar{y})\right]^2}{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}} \rightarrow \quad -1 \leq R \leq 1$$

$\bar{y} \rightarrow$ promedio de los valores y_i $\quad\bar{x} \rightarrow$ promedio de los valores x_i

- En el caso lineal R^2 es la correlación (mide cuan sensible es el valor de *y* respecto de lo que le pasa a *x*)
- Solo depende de los datos medidos x_i e y_i.
- Para calcularlo no necesito ningún ajuste de los datos.

Coeficiente de determinación R^2 (R-Squared): representa la porción de variación en las variables de respuesta que es explicada por los distintos valores de entrada.

$$\boxed{R^2 = \frac{\left[\sum(x_i - \bar{x}) \sum(y_i - \bar{y})\right]^2}{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \rightarrow \quad 0 \leq R^2 \leq 1$$

$\bar{y} \rightarrow$ promedio de los valores y_i $\bar{x} \rightarrow$ promedio de los valores x_i

- En el caso lineal R^2 es la correlación (mide cuan sensible es el valor de *y* respecto de lo que le pasa a *x*)

- Solo depende de los datos medidos x_i e y_i.

- Para calcularlo no necesito ningún ajuste de los datos.

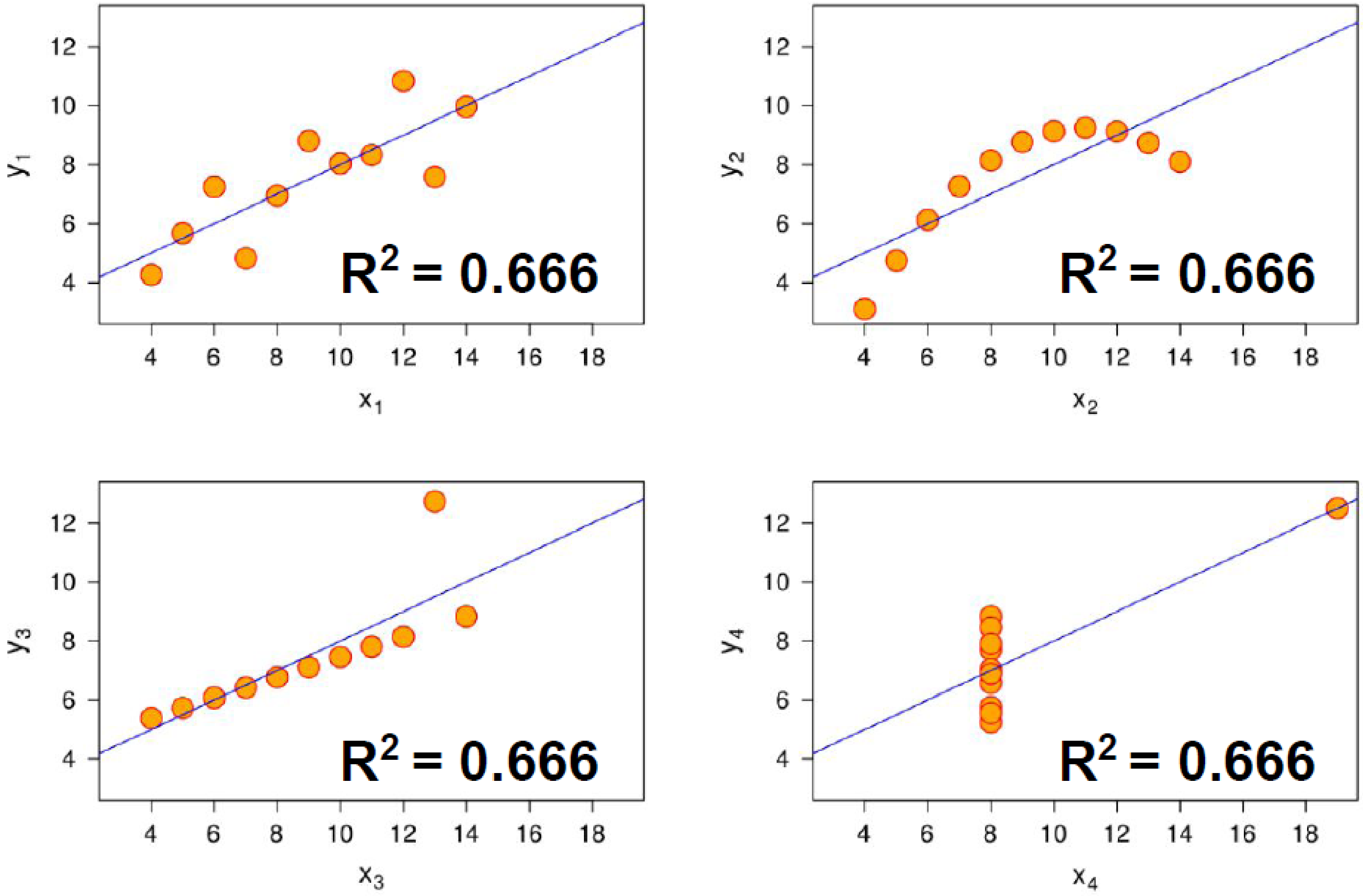
- No es necesariamente una medida de que tan bueno es un ajuste (ej: pendiente = 0).

Si hacemos un ajuste por cuadrados mínimos de una recta y obtengo m y b (y=m.x + b), R^2 está dando la correlación entre los datos.

$R \approx 1$ $\rightarrow$ los datos están muy correlacionados

$R = 0$ $\rightarrow$ los datos están completamente descorrelacionados (si varío el valor de *x*, a *y* no le importa)

Cuarteto de Anscombe



- 4 conjuntos de datos con las mismas propiedades estadísticas pero con gráficos diferentes (publicado en 1973 por el estadístico Francis John Anscombe).
- Resalta la importancia de mirar gráficamente un conjunto de datos antes de analizarlos.

¿Qué pasa si el modelo no es lineal?

¿Qué pasa si el modelo no es lineal?

Evaluar si se puede transformar la función en una función lineal (**linealización**)

¿Por qué linealizamos?

¿Criterios?

¿Qué pasa si el modelo no es lineal?

$\downarrow$

Evaluar si se puede transformar la función en una función lineal (**linealización**)

¿Por qué linealizamos? $\rightarrow$ Los modelos lineales son más fáciles de analizar.

$\downarrow$

Puedo aplicar regresión lineal por cuadrados mínimos.

¿Criterios? $\rightarrow$

- En base a consideraciones teóricas.
- A partir del gráfico de los datos podría surgir la necesidad de reexpresar las variables del modelo.

¿Qué pasa si el modelo no es lineal?

↓

Evaluar si se puede transformar la función en una función lineal (**linealización**)

¿Por qué linealizamos? → Los modelos lineales son más fáciles de analizar.

↓

Puedo aplicar regresión lineal por cuadrados mínimos.

¿Criterios? →

- En base a consideraciones teóricas.
- A partir del gráfico de los datos podría surgir la necesidad de reexpresar las variables del modelo.

Cuando linealizamos definimos nuevas variables.

¿Cómo linealizo? →

- Reemplazar las variables del modelo por nuevas variables → **clase de hoy**
- Graficar en escala logarítmica → **lo veremos en las próximas guías**

Ejemplo de linealización:

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

$\rightarrow$ función no lineal
(el gráfico T vs. L no es lineal)

Ejemplo de linealización:

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

$\rightarrow$ función no lineal
(el gráfico T vs. L no es lineal)

Si la reescribo: $$T = \frac{2\pi}{\sqrt{g}}\sqrt{L}$$

Vemos que se puede definir una nueva variable: $$u = \sqrt{L}$$

Ejemplo de linealización:

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

$\rightarrow$ función no lineal
(el gráfico T vs. L no es lineal)

Si la reescribo: $$T = \frac{2\pi}{\sqrt{g}}\sqrt{L}$$

Vemos que se puede definir una nueva variable: $$u = \sqrt{L}$$

Ahora T y u siguen una relación lineal:

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

Ejemplo de linealización:

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

$\rightarrow$ función no lineal (el gráfico T vs. L no es lineal)

Si la reescribo: $$T = \frac{2\pi}{\sqrt{g}}\sqrt{L}$$

Vemos que se puede definir una nueva variable: $$u = \sqrt{L}$$

Ahora T y u siguen una relación lineal:

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

Grafico T vs. u **(considerar las incertezas ΔT y Δu)**.

$\downarrow$

Puedo aplicar regresión lineal por cuadrados mínimos

Ejemplo de linealización:

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

$\rightarrow$ función no lineal
(el gráfico T vs. L no es lineal)

Si la reescribo: $$T = \frac{2\pi}{\sqrt{g}}\sqrt{L}$$

Vemos que se puede definir una nueva variable: $$u = \sqrt{L}$$

Ahora T y u siguen una relación lineal:

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

Grafico T vs. u **(considerar las incertezas ΔT y Δu)**.

$\downarrow$

Puedo aplicar regresión lineal por cuadrados mínimos

Si elevo al cuadrado la expresión de arriba:

$$T^2 = \frac{4\pi^2}{g}L$$

Vemos que se puede definir una nueva variable: $$v = T^2$$

Ejemplo de linealización:

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

→ función no lineal (el gráfico T vs. L no es lineal)

Si la reescribo: $$T = \frac{2\pi}{\sqrt{g}}\sqrt{L}$$

Vemos que se puede definir una nueva variable: $$u = \sqrt{L}$$

Ahora T y u siguen una relación lineal:

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

Grafico T vs. u **(considerar las incertezas $\boldsymbol{\Delta T}$ y $\boldsymbol{\Delta u}$)**.

↓

Puedo aplicar regresión lineal por cuadrados mínimos

Si elevo al cuadrado la expresión de arriba:

$$T^2 = \frac{4\pi^2}{g}L$$

Vemos que se puede definir una nueva variable: $$v = T^2$$

Ahora L y v siguen una relación lineal:

$$v = \frac{4\pi^2}{g}L$$

Grafico L vs. v **(considerar las incertezas $\boldsymbol{\Delta L}$ y $\boldsymbol{\Delta v}$)**.

↓

Puedo aplicar regresión lineal por cuadrados mínimos

Supongamos que quiero graficar: ***T* vs. *u***

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

T	$Err(T)$	$\sqrt{L}$	$Err(\sqrt{L})$
s	s	$\mathrm{m}^{1/2}$	$\mathrm{m}^{1/2}$

¿Cómo calculo la incerteza Δu?

Gráfico *T* vs. *u*:

¿Cuál elijo como variable independiente (x) y cuál como variable dependiente (y)?

Supongamos que quiero graficar: ***T* vs. *u***

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

T	$Err(T)$	$\sqrt{L}$	$Err(\sqrt{L})$
s	s	$\mathrm{m}^{1/2}$	$\mathrm{m}^{1/2}$

¿Cómo calculo la incerteza Δu? $\rightarrow$ Propagación de errores

Gráfico *T* vs. *u*:

¿Cuál elijo como variable independiente (x) y cuál como variable dependiente (y)?

Eje x $\rightarrow$ variable medida con mayor precisión (menor error relativo ó error relativo porcentual).

¿Por qué?

Supongamos que quiero graficar: ***T* vs. *u***

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

T	$Err(T)$	$\sqrt{L}$	$Err(\sqrt{L})$
s	s	$\mathrm{m}^{1/2}$	$\mathrm{m}^{1/2}$

¿Cómo calculo la incerteza Δ*u*? → Propagación de errores

Gráfico *T* vs. *u*:

¿Cuál elijo como variable independiente (*x*) y cuál como variable dependiente (*y*)?

Eje *x* → variable medida con mayor precisión (menor error relativo ó error relativo porcentual).

¿Por qué? → Cuadrados mínimos ponderados tiene en cuenta solamente el error en el eje *y*.

Comparo errores relativos → $\frac{\Delta u}{u} \quad vs. \quad \frac{\Delta T}{T}$

Supongamos que quiero graficar: ***T* vs. *u***

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

T	$Err(T)$	$\sqrt{L}$	$Err(\sqrt{L})$
s	s	$\mathrm{m}^{1/2}$	$\mathrm{m}^{1/2}$

¿Cómo calculo la incerteza Δu? $\rightarrow$ Propagación de errores

Gráfico *T* vs. *u*:

Si grafico T en función de u $\rightarrow$ pendiente = $\dfrac{2\pi}{\sqrt{g}}$

Si grafico u en función de T $\rightarrow$ pendiente = $\dfrac{\sqrt{g}}{2\pi}$

Gráfico del período de un péndulo simple en función del largo de la soga.

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

¿Puedo hacer una regresión lineal por cuadrados mínimos?

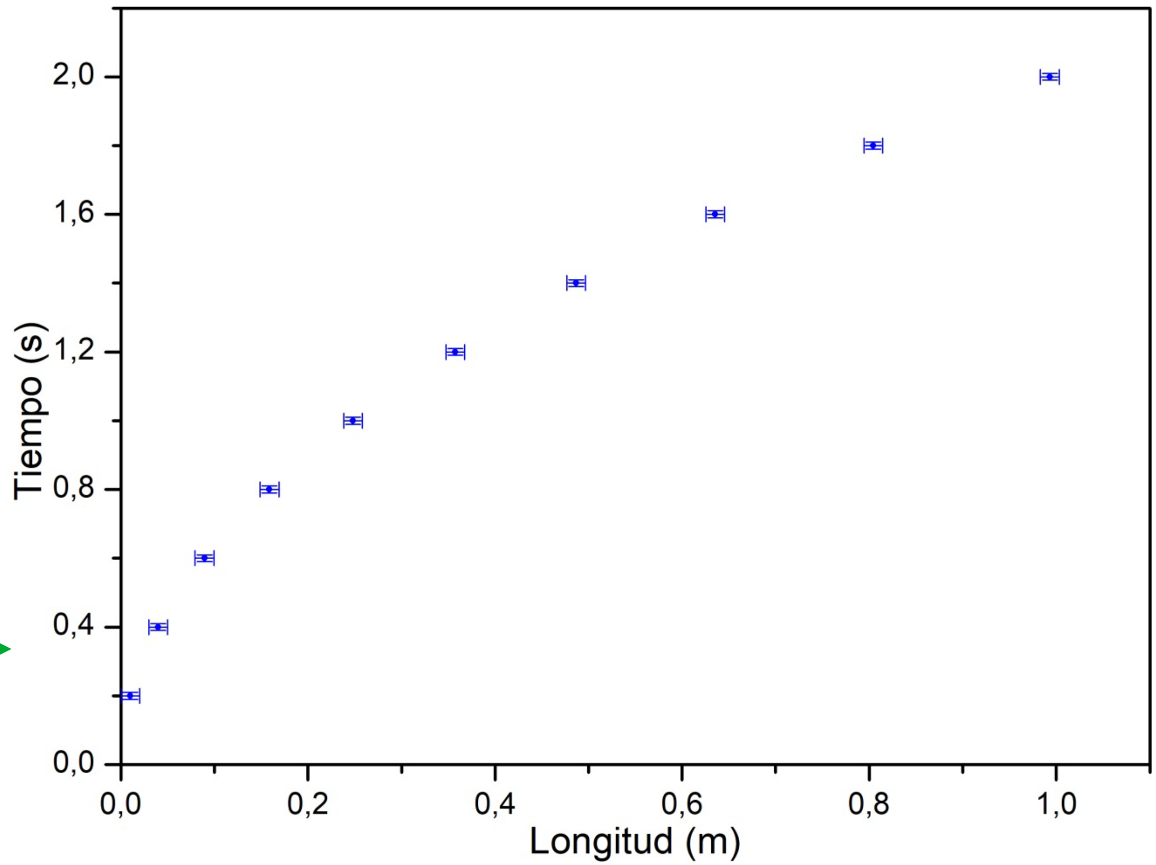
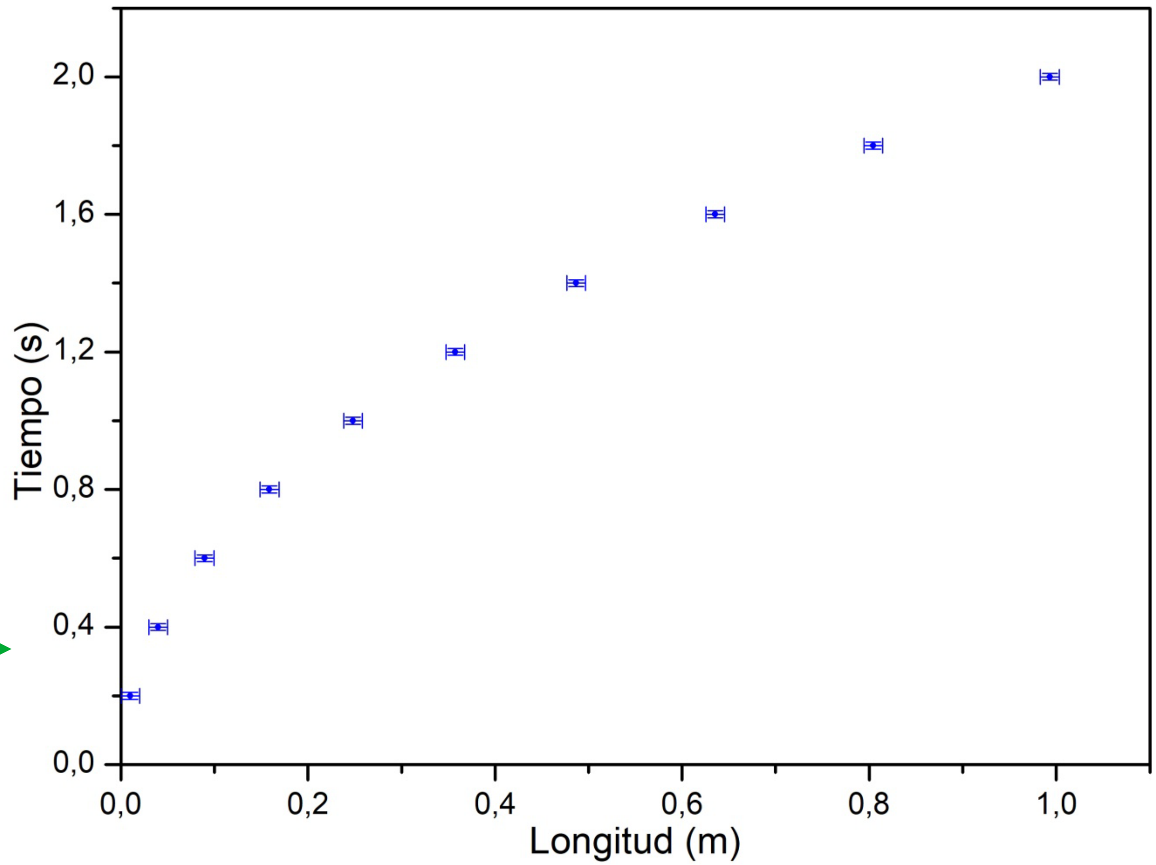


Gráfico del período de un péndulo simple en función del largo de la soga.

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



¿Puedo hacer una regresión lineal por cuadrados mínimos?

NO. Conocemos la física del problema: sabemos que en un péndulo simple la **relación entre T y L no es lineal** (ver ecuación de arriba).