Laboratorio 1

Medición de la aceleración de la gravedad
Regresión lineal

Mónica Agüero – 2C 2025

¿Cómo podemos obtener experimentalmente el valor de una magnitud física (MF)?

1- SIEMPRE hay que **buscar las LEYES FÍSICAS que conozcamos QUE CONTENGAN LA MF que deseamos calcular.**

2- Buscar CUÁL O **CUÁLES PUEDEN APLICARSE con el equipamiento con el que contamos.**

En muchos experimentos vamos a medir valores de dos variables físicas diferentes. Por ejemplo:

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

Si medimos T y L de forma directa o indirecta podemos obtener g de manera indirecta.

¿Cómo podemos obtener experimentalmente el valor de una magnitud física (MF)?

1- SIEMPRE hay que **buscar las LEYES FÍSICAS que conozcamos QUE CONTENGAN LA MF que deseamos calcular.**

2- Buscar CUÁL O **CUÁLES PUEDEN APLICARSE con el equipamiento con el que contamos.**

En muchos experimentos vamos a medir valores de dos variables físicas diferentes. Por ejemplo:

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

Si medimos *T* y *L* de forma directa o indirecta podemos obtener *g* de manera indirecta.

¿Qué pasa si tenemos muchos valores de *T* y de *L* de otro péndulo con diferente longitud?

$$T_1 = T_{01} \pm \Delta T_1 \qquad l_1 = l_{01} \pm \Delta l_1 \quad \rightarrow \quad g_1 = g_{01} \pm \Delta g_1$$

$$T_2 = T_{02} \pm \Delta T_2 \qquad l_2 = l_{02} \pm \Delta l_2 \quad \rightarrow \quad g_2 = g_{02} \pm \Delta g_2$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$T_n = T_{0n} \pm \Delta T_n \qquad l_n = l_{0n} \pm \Delta l_n \quad \rightarrow \quad g_n = g_{0n} \pm \Delta g_n$$

Podemos pensar en calcular muchos valores de *g* y promediarlos. ¿Pero esto es lo mejor? ¿Cuál sería la ventaja? ¿Qué pasa si tenemos otro tipo de relaciones?

Objetivos de la clase de hoy

- Determinar la aceleración de la gravedad *g* a partir de un experimento de péndulo simple de longitud variable.

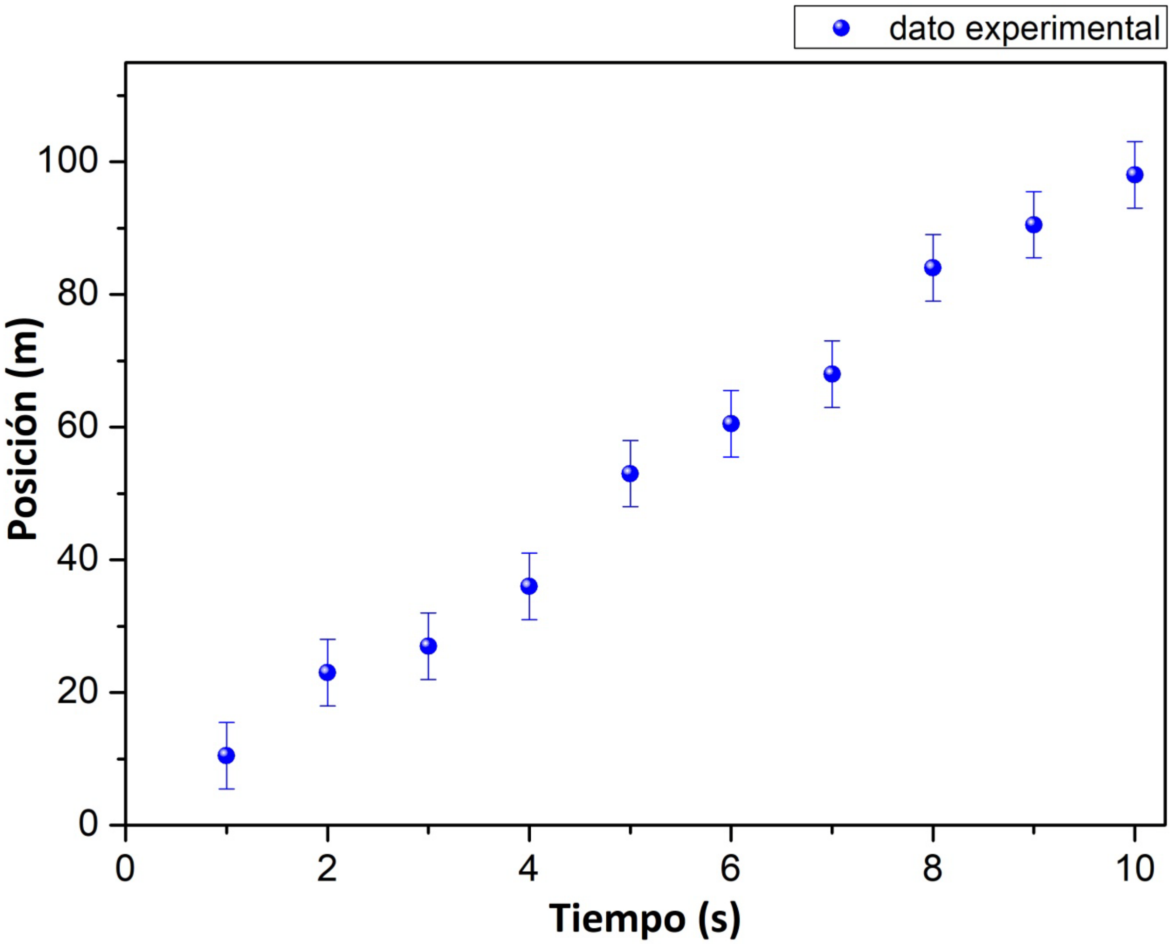
$$g = (g_0 \pm \Delta g) unidad$$

- Obtener el resultado de una MF a partir de un NUEVO MÉTODO de análisis.

- Aprender a utilizar un nuevo instrumento de medición.

En muchos experimentos vamos a medir varios valores de dos variables físicas diferentes para investigar la relación matemática que hay entre ellas y obtener el valor de una MF.

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

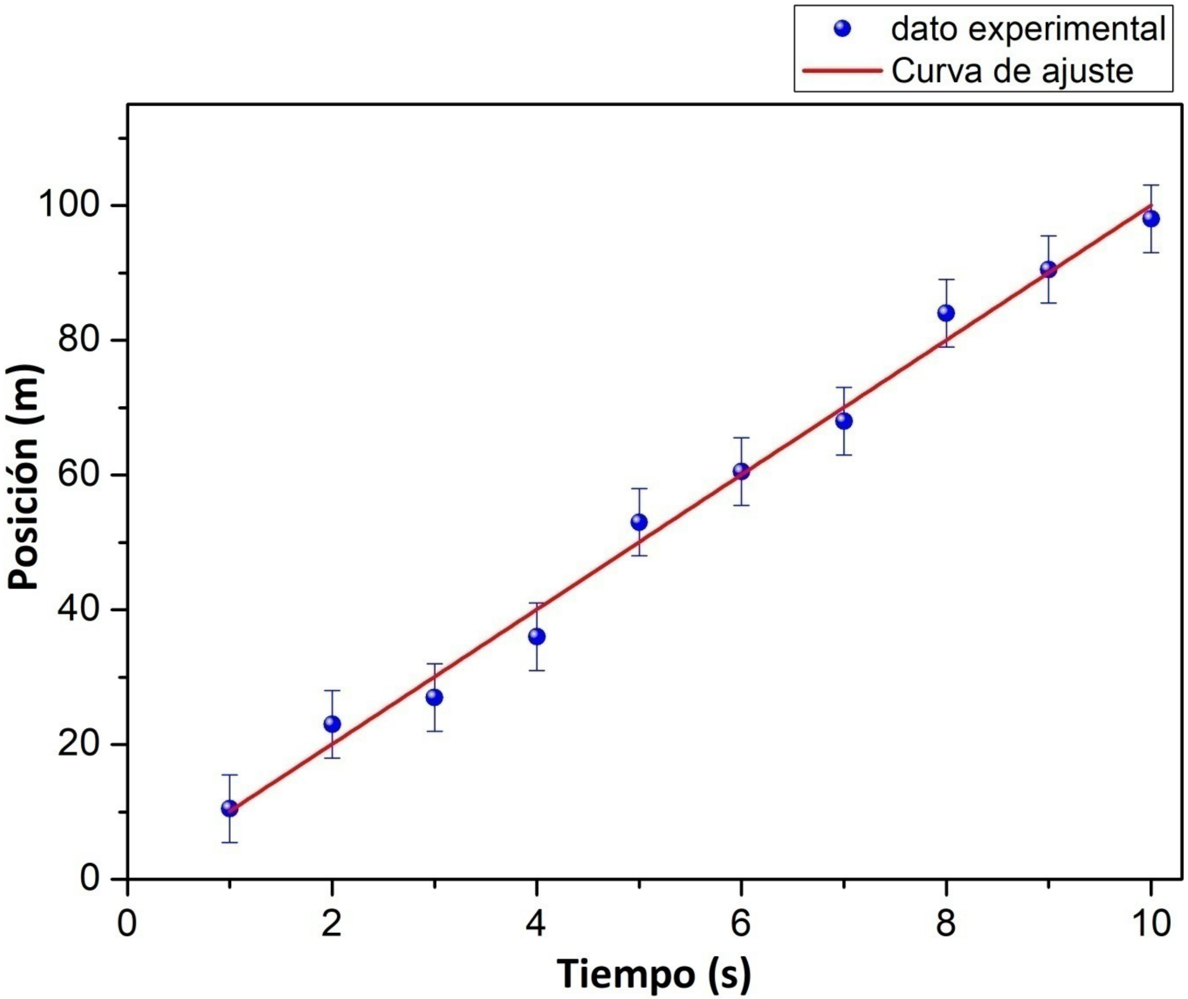
Los datos parecen seguir una relación lineal.

Supongamos que tenemos un modelo que vincula dos magnitudes x e y.

$$y = f(x)$$

Modelo lineal:

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



Mediciones:

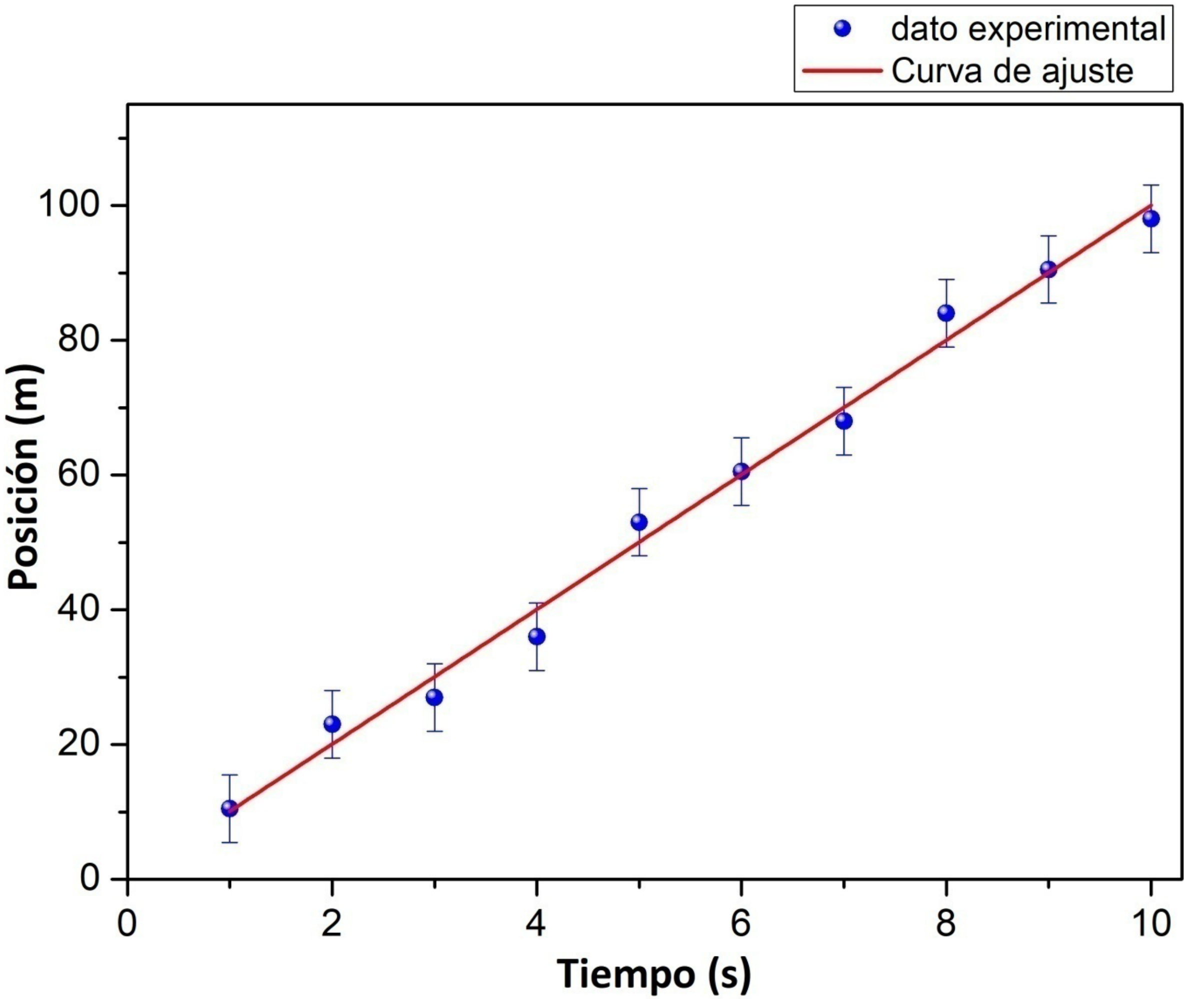
$$(x_i \pm \Delta x_i, y_i \pm \Delta y_i)$$

$x \rightarrow$ 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**.

Hipótesis:

- Los datos deben ser independientes.
- Se asume que los errores de cada medida tienen una distribución normal.

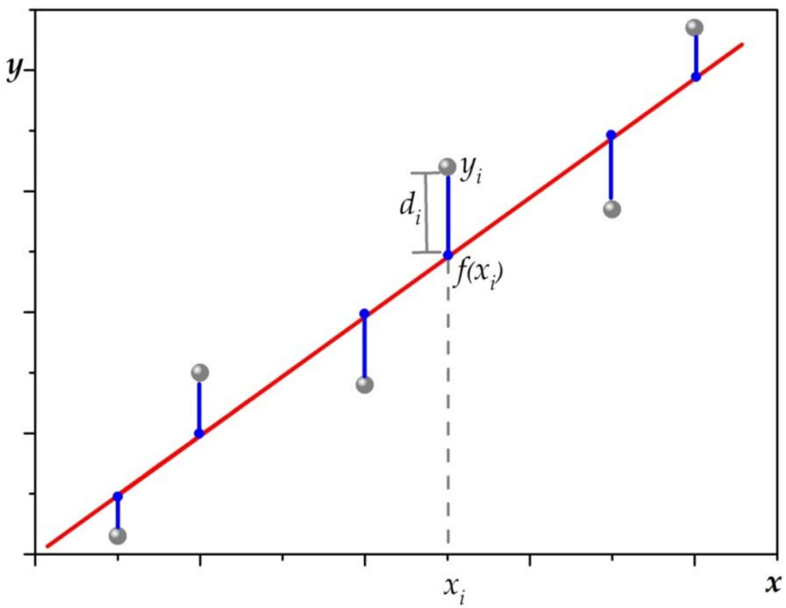
Nos permite medir el valor de los parámetros de nuestro modelo a partir de mediciones de las variables.

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**.

Hipótesis:
- Los datos deben ser independientes.
- Se asume que los errores de cada medida tienen una distribución normal.

Nos permite medir el valor de los parámetros de nuestro modelo a partir de mediciones de las variables.

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

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

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**.

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

$$M = \sum_{i=1}^{N} (d_i)^2 = \sum_{i=1}^{N} (y_i - A.x_i - B)^2 \quad \rightarrow \quad M \text{ sea un mínimo}$$

(minimizamos la suma de los residuos d_i al cuadrado) $\quad \dfrac{\partial M}{\partial A} = 0 \qquad \dfrac{\partial M}{\partial B} = 0$

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 de los parámetros A y B $\rightarrow$

Si los parámetros son números que yo puedo calcular a partir de mis mediciones, **¡los parámetros en sí son mediciones indirectas!** $\Rightarrow$ Los parámetros tienen incertezas y las puedo calcular usando **propagación de errores**.

Incertezas de los parámetros A y $B \rightarrow$

Necesito la incerteza de los puntos pero Cuadrados Mínimos Ordinarios no tiene en cuenta las incertezas. Como tengo un modelo que supuestamente describe el "valor verdadero" de mis mediciones, a falta de una mejor aproximación, asumo que la incerteza en cada punto es el residuo:

$$\Delta y_i = d_i = y_i - (Ax_i + B)$$

Propagando errores $\rightarrow$

$$\sigma_A = \sigma_y \sqrt{\frac{N}{N \sum x_i^2 - (\sum x_i)^2}}$$

$$\sigma_B = \sigma_y \sqrt{\frac{\sum x_i^2}{N \sum x_i^2 - (\sum x_i)^2}}$$

siendo $$\sigma_y = \sqrt{\frac{1}{N-2} \sum (y_i - Ax_i - B)}$$

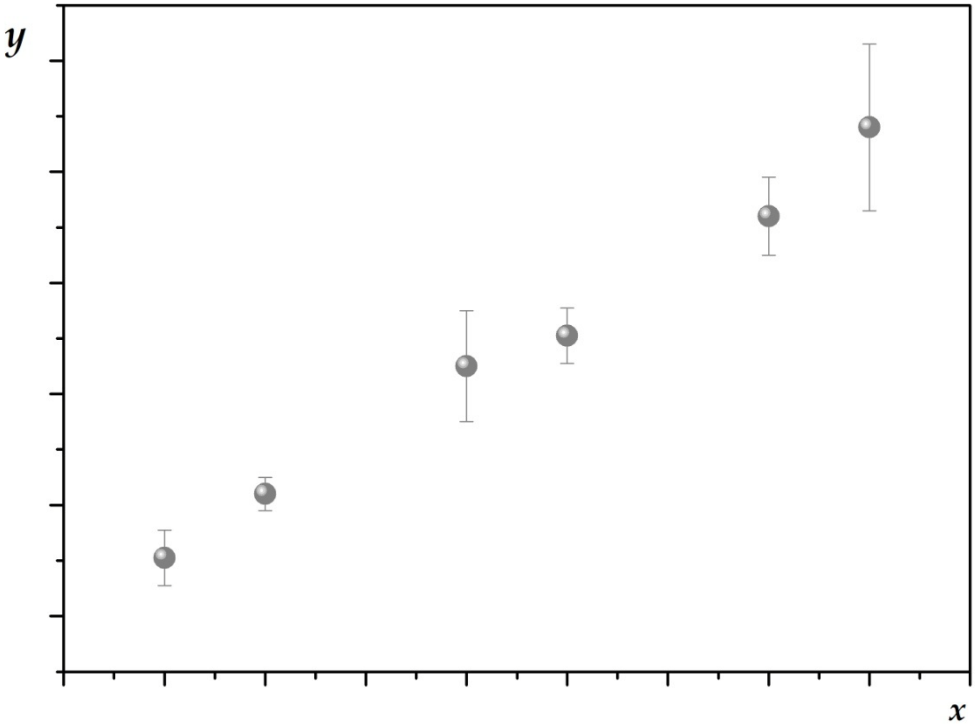
Detalles: D.C. Baird, *Experimentación*, p. 128 y p.172.

Pendiente $\rightarrow$ $A \pm \sigma_A$

Ordenada al origen $\rightarrow$ $B \pm \sigma_B$

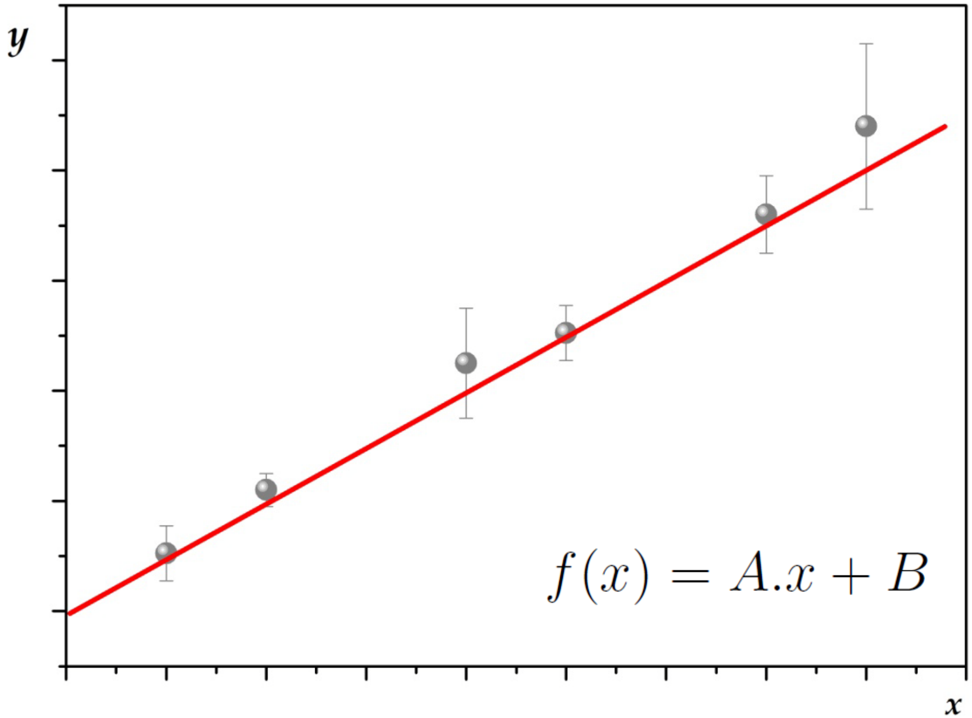
¿Todos los puntos son equivalentes? ¿Confiamos más en alguno que en otro?

Regresión lineal por cuadrados mínimos <u>ponderados</u>



¿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 → se asigna mayor importancia a los d_i provenientes de valores de los y_i que sean más precisos.

→

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

En vez de los residuos al cuadrado, minimizamos algo que tenga en cuenta las incertezas:

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{d_i}{\Delta y_i}\right)^2 = \sum_{i=1}^{N} \left(\frac{y_i - A.x_i - B}{\Delta y_i}\right)^2 \xrightarrow{\text{minimizar}}$$

Se obtienen expresiones para los parámetros *A* y *B*, y sus incertezas σ_A y σ_B del modelo.

Definimos las cantidades w_i y Δ:

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

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

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

Regresión lineal por **cuadrados mínimos ponderados**.

Incertezas:

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

Pendiente $\rightarrow A \pm \sigma_A$

Ordenada al origen $\rightarrow B \pm \sigma_B$

¿Cómo sé si el modelo es adecuado? → **Parámetros de BONDAD**

Los *parámetros de Bondad* pueden darnos una idea de la **discrepancia entre los valores observados (datos experimentales) y los esperados según el modelo de estudio.**

¿Cómo sé si el modelo es adecuado? ➡ **Parámetros de BONDAD**

Los *parámetros de Bondad* pueden darnos una idea de la **discrepancia entre los valores observados (datos experimentales) y los esperados según el modelo de estudio.**

Coeficiente de Correlación de Pearson (r)

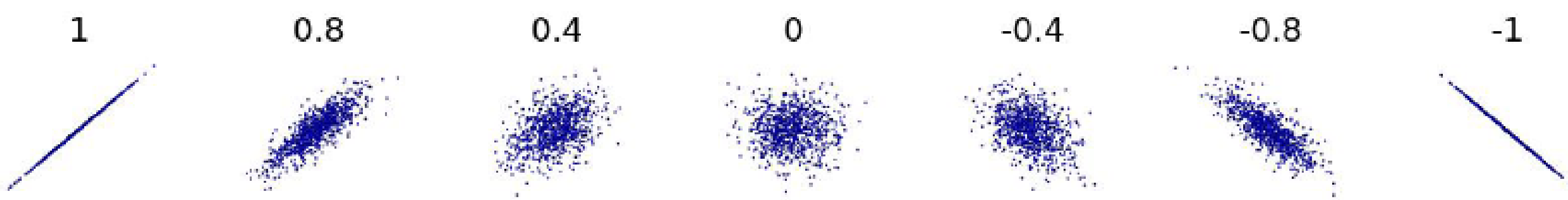
Indica cuán fuerte es la correlación entre las variables x e y
(mide cuan sensible es el valor de y respecto de lo que le pasa a x)

$$\boxed{r = \frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}} = \frac{S_{xy}}{S_x S_y} \qquad -1 \leq r \leq 1}$$

$$Var(x) = S_x^2 = \frac{1}{N}\sum(x_i - \bar{x})^2$$

$$Var(y) = S_y^2 = \frac{1}{N}\sum(y_i - \bar{y})^2$$

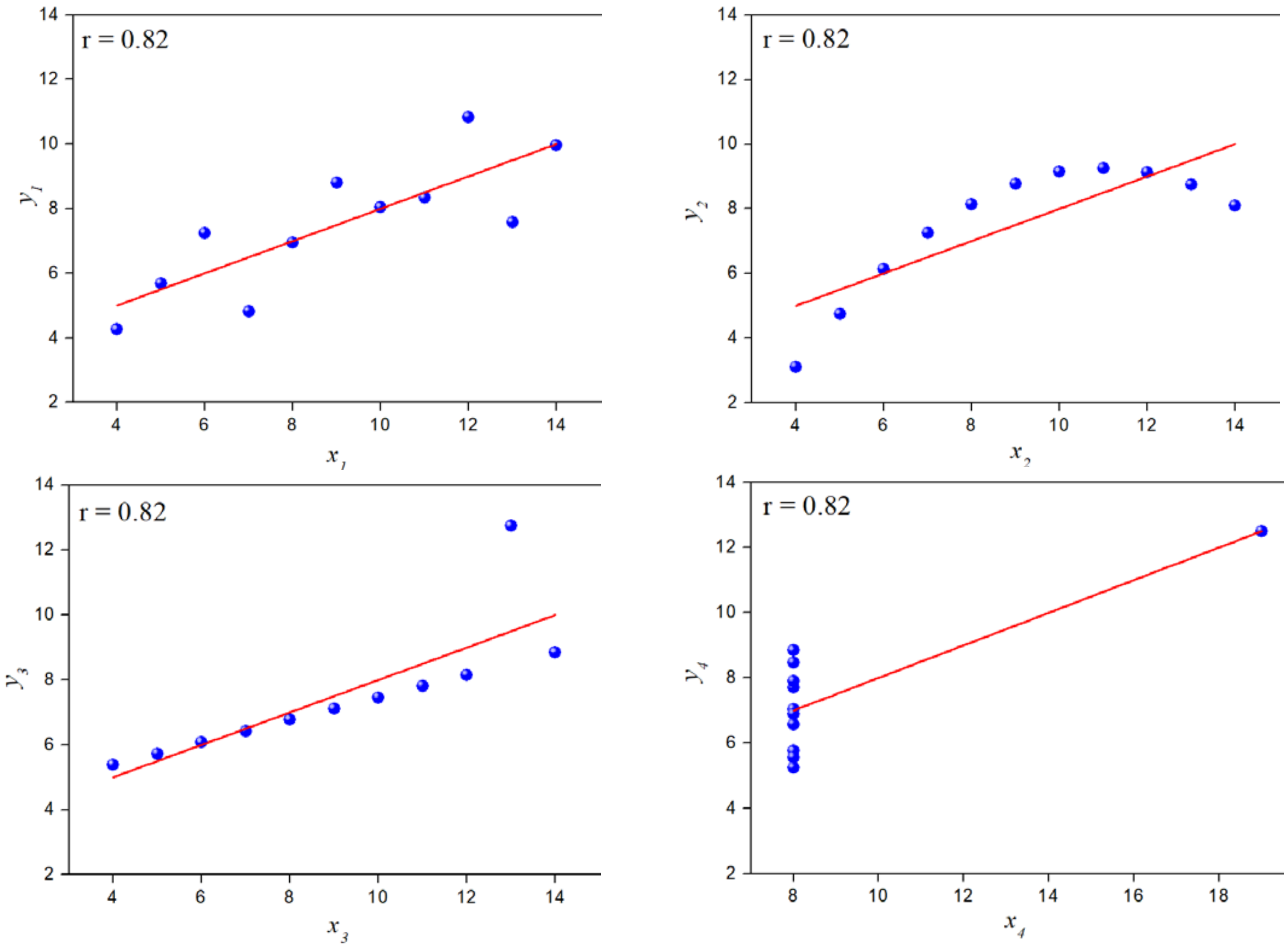
$$Cov(x,y) = S_{xy} = \frac{1}{N}\sum(x_i - \bar{x})(y_i - \bar{y})$$



- $|r| \approx 1$ $\rightarrow$ los datos están muy correlacionados.
- $r = 0$ $\rightarrow$ los datos están completamente descorrelacionados.
- Sólo depende de los datos medidos x_i e y_i. Para calcularlo no necesito hacer ningún ajuste de los datos.

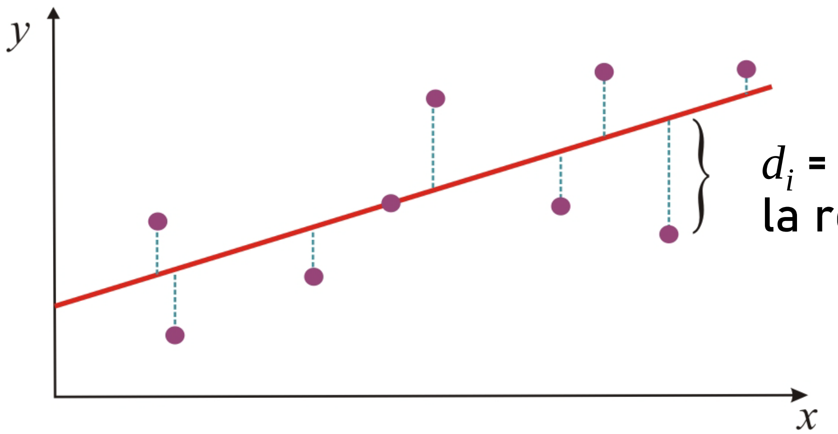
Cuarteto de Anscombe

¡Notar que estos casos tienen igual valor de *r*!

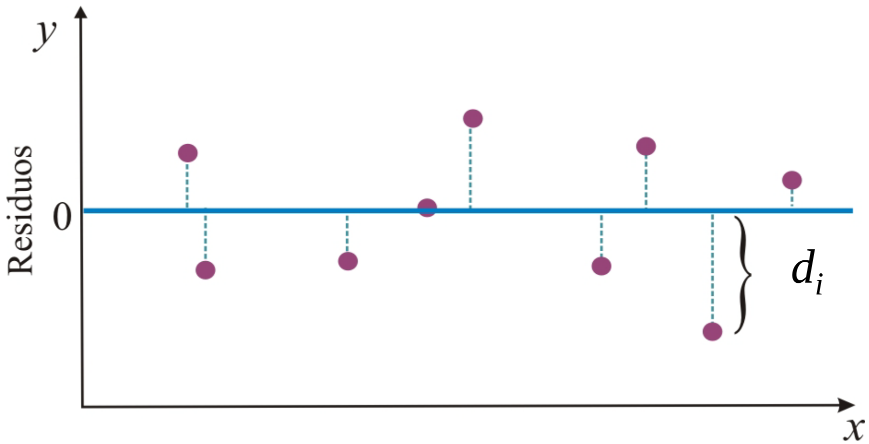


Por eso la IMPORTANCIA de mirar gráficamente un conjunto de datos antes de analizarlos.

Gráfico de Residuos



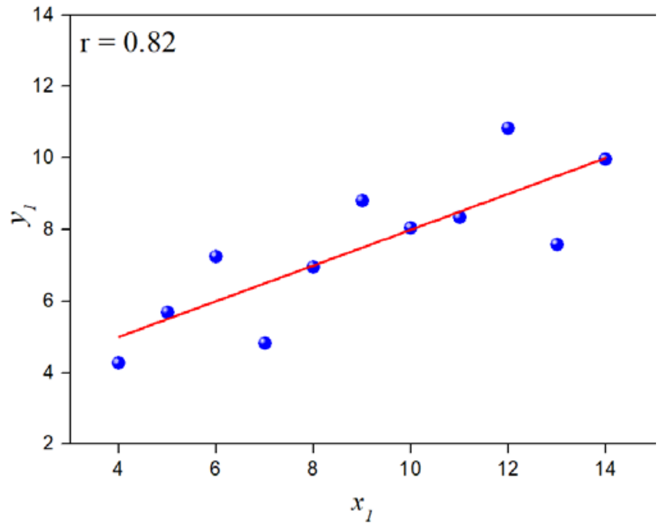
d_i = distancia de los puntos experimentales a la recta, en "y"



Los residuos deben estar distribuidos en forma aleatoria alrededor del cero.

NO DEBEN tener ESTRUCTURA

Analicemos el gráfico de residuos para algunos casos del Cuarteto de Anscombe:



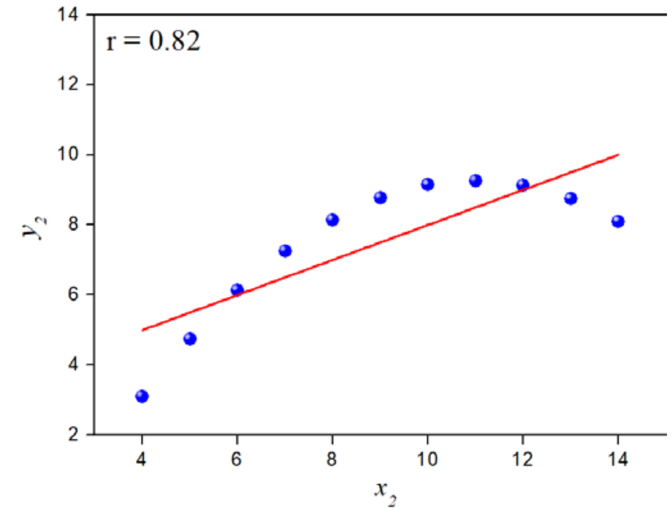
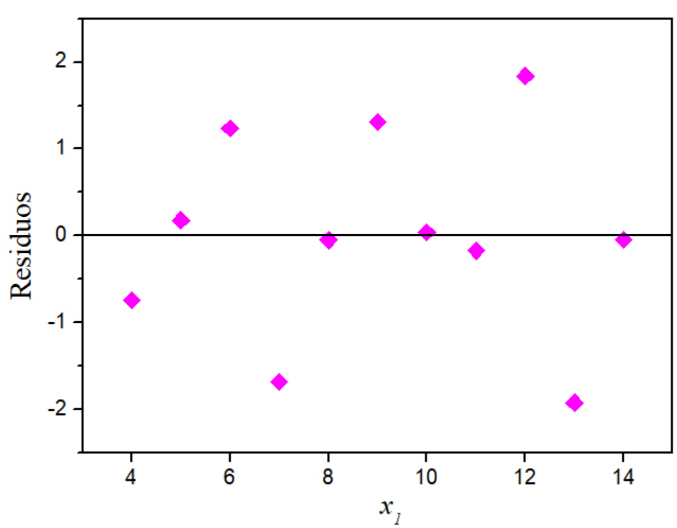
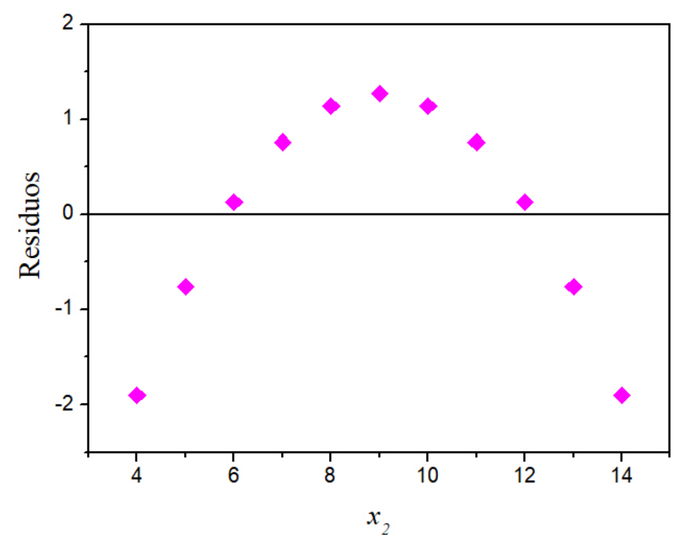


Gráfico de residuos (Residuo = valor observado – valor de la predicción)





Muestra un comportamiento aleatorio de los residuos.

Chi cuadrado reducido χ^2_ν

Dimensiona cuánto difieren los datos experimentales de los del modelo.
Pesa fuertemente la incerteza Δy.

$$\chi^2 = \sum \left[\frac{y_i - (A.x + B)}{\Delta y_i} \right]^2$$

Caso lineal:
Chi cuadrado reducido ➡

$$\chi^2_\nu = \frac{\chi^2}{N-2}$$

$N \rightarrow$ número de datos
$2 \rightarrow$ Cantidad de parámetros
$\nu = N - 2 \rightarrow$ los grado de libertad

Chi cuadrado reducido χ^2_ν

Dimensiona cuánto difieren los datos experimentales de los del modelo. Pesa fuertemente la incerteza Δy.

$$\chi^2 = \sum \left[\frac{y_i - (A.x + B)}{\Delta y_i} \right]^2$$

Caso lineal: Chi cuadrado reducido ⇒

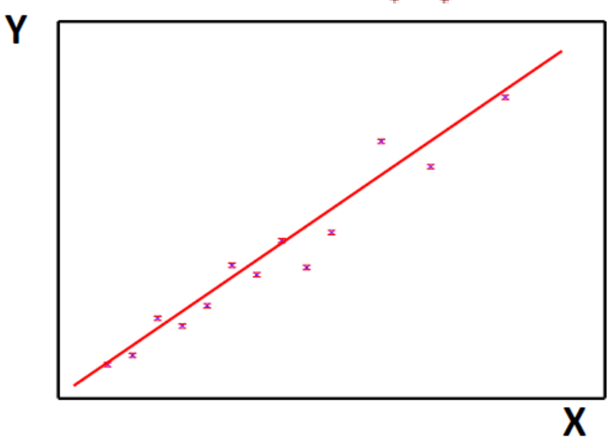
$$\chi^2_\nu = \frac{\chi^2}{N-2}$$

$N \rightarrow$ número de datos

$2 \rightarrow$ Cantidad de parámetros

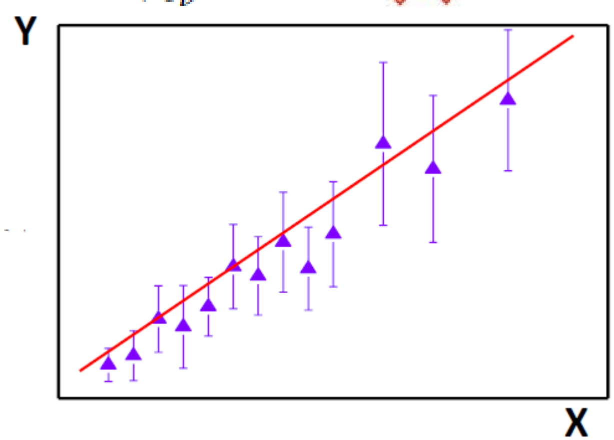
ν= N - 2 $\rightarrow$ los grado de libertad

$\chi^2_\nu \gg 1$



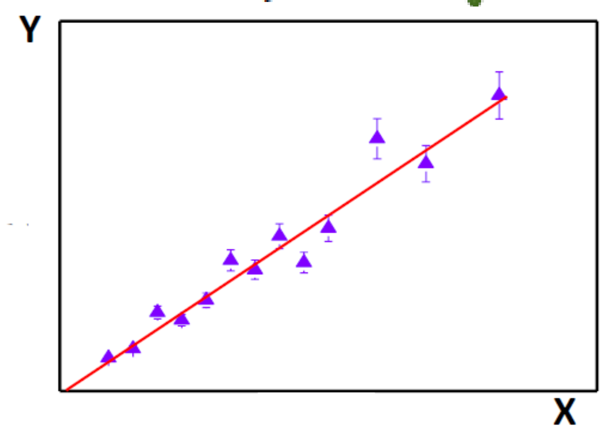
Incertezas subestimadas (suponiendo que la teoría es correcta). Si la incerteza está bien estimada entonces el modelo no es adecuado.

$\chi^2_\nu \ll 1$



El modelo es presumiblemente compatible con el experimento pero las incertezas podrían estar sobreestimadas.

$\chi^2_\nu \sim 1$



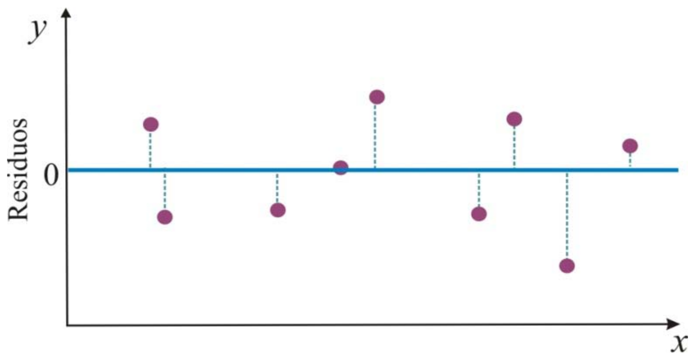
El modelo es presumiblemente compatible con el experimento.

Parámetros de BONDAD: resumen

Si el modelo es adecuado para describir mis datos experimentales, esperamos:

- **Coeficiente de Correlación de Pearson** → $|r| \approx 1$

- **Gráfico de residuos** → Sin ESTRUCTURA



- **Chi-cuadrado reducido** → $\chi^2_\nu \sim 1$

¿Qué pasa si el modelo no es lineal?

¿Qué pasa si el modelo no es lineal?

Evalúar 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.

¿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 $\rightarrow$ **clase de hoy**
- Graficar en escala logarítmica $\rightarrow$ **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: $T = \dfrac{2\pi}{\sqrt{g}}\sqrt{L}$

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

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:

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

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:

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

→ 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 ΔT y Δu).**

↓

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 ΔL y Δv).**

↓

Puedo aplicar regresión lineal por cuadrados mínimos

Supongamos que quiero graficar: *L* **vs.** v y aplicar regresión lineal por cuadrados mínimos ponderados.

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

L	Error L	v	Error v

¿Cómo calculo la incerteza Δv?

Gráfico de las variables *L* y v:

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

Supongamos que quiero graficar: *L* **vs.** v y aplicar regresión lineal por cuadrados mínimos ponderados.

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

L	Error L	v	Error v

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

Gráfico de las variables *L y* v:

¿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: *L* **vs.** v y aplicar regresión lineal por cuadrados mínimos ponderados.

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

L	Error L	v	Error v

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

Gráfico de las variables *L* y v:

¿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? $\rightarrow$ Cuadrados mínimos ponderados tiene en cuenta solamente el error en el eje y.

Comparo errores relativos $\rightarrow$ $\dfrac{\Delta v}{v}$ $vs.$ $\dfrac{\Delta L}{L}$

Supongamos que quiero graficar: L vs. v y aplicar regresión lineal por cuadrados mínimos ponderados.

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

L	Error L	v	Error v

¿Cómo calculo la incerteza Δv? → Propagación de errores

Gráfico de las variables L y v:

Si grafico L (eje y) y v (eje x) → pendiente = $\dfrac{g}{4\pi^2}$

Si grafico v (eje y) y L (eje x) → pendiente = $\dfrac{4\pi^2}{g}$