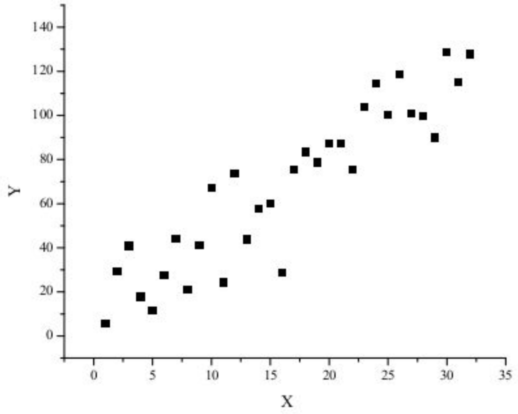
Cálculo de la aceleración de la gravedad: método de cuadrados mínimos

Laboratorio 1
Departamento de física -FCEyN- UBA

Objetivo: estudiar la dependencia funcional entre dos magnitudes y utilizar modelos que mejor describan* nuestros datos experimentales

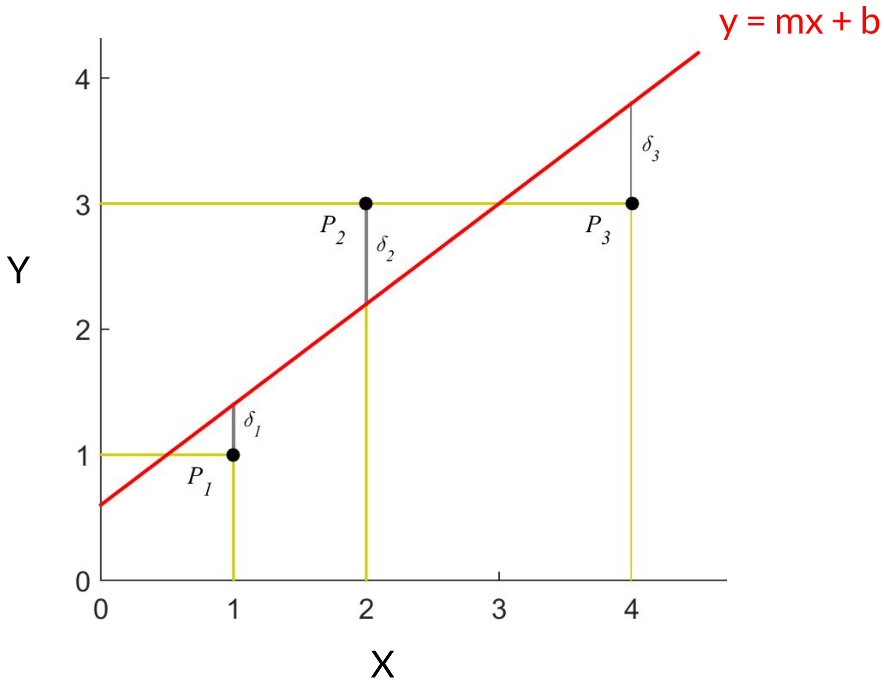


- Supongamos que medimos de forma independiente las magnitudes X e Y, tal que su conjunto de mediciones se describa como: X_i i= 1,...,N y Y_i i= 1,...,N

- Vamos a analizar el caso más simple: la dependencia lineal
- Para ello, asumimos un cierto modelo teórico: $y = mx + b$
- ¿Cómo encuentro los parámetros del modelo (m y b) que mejor ajustan los datos experimentales?¿Cuál es la mejor recta?

* Se dice también "que mejor se ajustan a" y de ahí viene hablar de "ajuste"

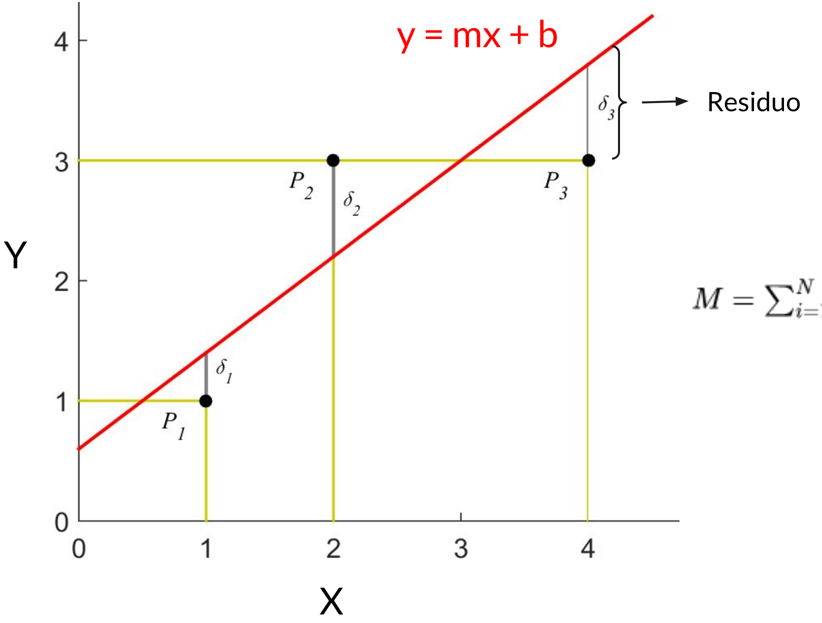
Método de cuadrados mínimos



- Buscamos una recta que pase lo más cerca posible de la distribución de puntos (nuestros datos experimentales)
- Para ello, esta recta debe minimizar las distancias cuadráticas entre los puntos y la recta.
- Consideremos el caso más sencillo:

$$\boxed{\frac{\Delta x}{x} < \frac{\Delta y}{y}}$$

Método de cuadrados mínimos



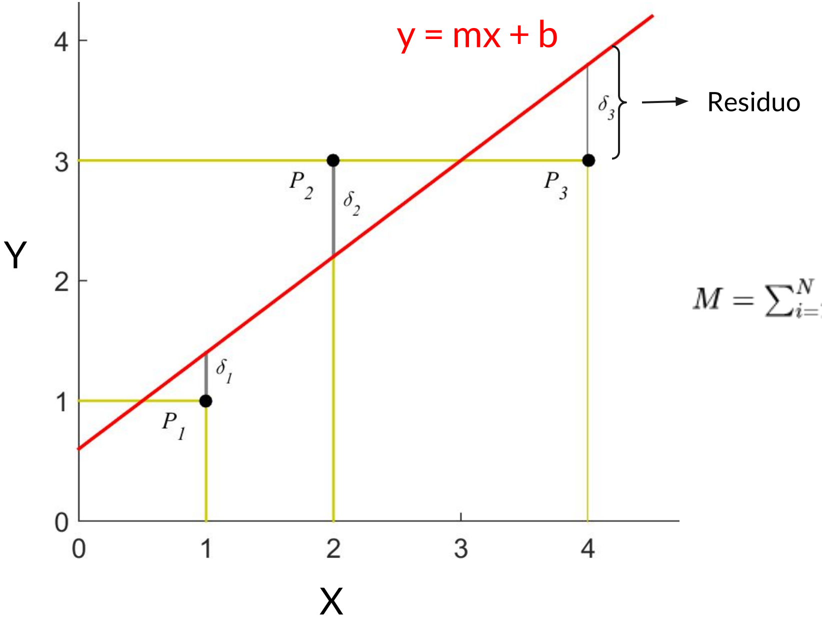
Sea la recta y = mx+b, vamos a querer minimizar:

$$(\delta y_i)^2 = [y_i - (mx_i + b)]^2$$

Esto se conoce como el residuo. Sumando los residuos:

$$M = \sum_{i=1}^{N} (\delta y_i)^2 = \sum_{i=1}^{N} y_i^2 + m \sum_{i=1}^{N} x_i^2 + Nb^2 + 2mb \sum_{i=1}^{N} x_i - 2m \sum_{i=1}^{N} x_i y_i - 2b \sum_{i=1}^{N} y_i$$

Método de cuadrados mínimos



Sea la recta y = mx+b, vamos a querer minimizar:

$$(\delta y_i)^2 = [y_i - (mx_i + b)]^2$$

Esto se conoce como el residuo. Sumando los residuos:

$$M = \sum_{i=1}^{N} (\delta y_i)^2 = \sum_{i=1}^{N} y_i^2 + m \sum_{i=1}^{N} x_i^2 + Nb^2 + 2mb \sum_{i=1}^{N} x_i - 2m \sum_{i=1}^{N} x_i y_i - 2b \sum_{i=1}^{N} y_i$$

Buscamos encontrar los parámetros m y b que minimicen la distancia de los datos al modelo. La condición de mínimo está dada por:

$$\frac{\partial M}{\partial m} = 0 \quad \text{y} \quad \frac{\partial M}{\partial b} = 0$$

Método de cuadrados mínimos

$$\frac{\partial M}{\partial m} = 0 \longrightarrow 2m\sum_{i=1}^{N} x_i^2 + 2b\sum_{i=1}^{N} x_i - 2\sum_{i=1}^{N}(x_i y_i) = 0$$

$$\frac{\partial M}{\partial b} = 0 \longrightarrow 2Nb + 2m\sum_{i=1}^{N} x_i - 2\sum_{i=1}^{N} y_i = 0$$

Tenemos un sistema de 2 ecuaciones y 2 incógnitas: Despejo!

$$\boxed{m = \frac{N\sum_{i=1}^{N}(x_i y_i) - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}}$$

y

$$\boxed{b = \frac{\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N}(x_i y_i)}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}}$$

¿Cómo son los errores de m y b?

Esto es sólo válido cuando todas las medidas de **y** tienen igual precisión!

$$m = \frac{N \sum_{i=1}^{N} (x_i y_i) - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}$$

$$b = \frac{\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} (x_i y_i)}{N \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}$$

Propagando errores! → Tengo que propagar en estas ecuaciones el error de y_i

$$\Delta m = S_y \sqrt{\frac{N}{N \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}}$$

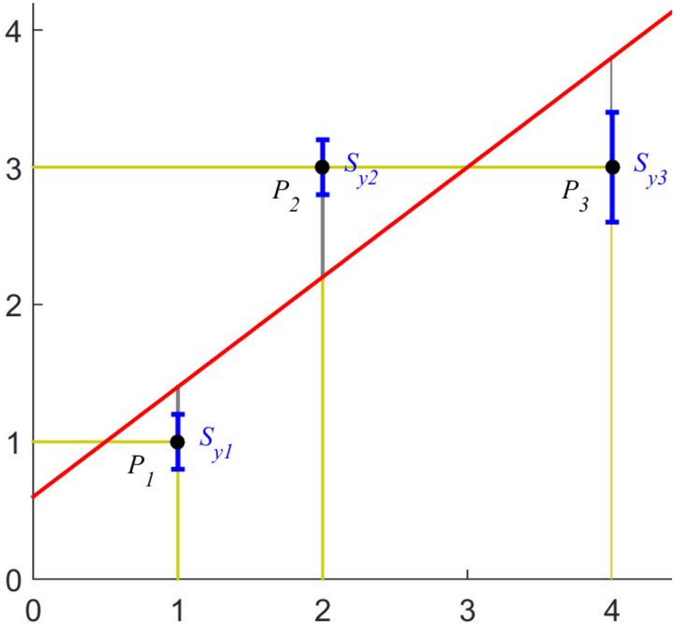
y

$$\Delta b = S_y \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}}$$

Con $S_y = \sqrt{\frac{\sum_{i=1}^{N} (\delta y_i)^2}{N-2}}$

Para más detalles de las cuentas ---> Ver D. C. Baird. Prentice Hall (1991). Apéndice 2

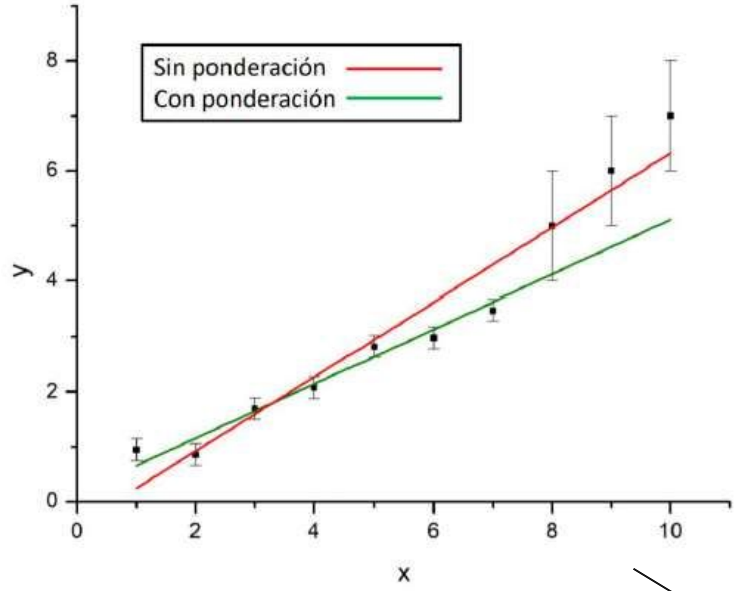
Cuadrados mínimos ponderados



Hasta ahora, sólo consideramos a los residuos como error, pero ¿Existe un método que tenga en consideración los errores de cada medición para obtener la recta? ¿Qué pasa si los puntos no pesan lo mismo?

En lugar de minimizar la suma de los residuos, minimizamos χ^2:

$$\chi^2 = \sum_{i=1}^{N} \left[\frac{(y_i - (mx_i + b))}{S_i} \right]^2$$



$$\chi^2 = \sum_{i=1}^{N} \left[\frac{(y_i - (mx_i + b))}{S_i} \right]^2 \longrightarrow$$ Ponderados

$$M = [y_i - (mx_i + b)]^2 \longrightarrow$$ No Ponderados

Son más relevantes las medidas con menor incerteza!

De forma análoga, calculamos las expresiones para m y b,

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

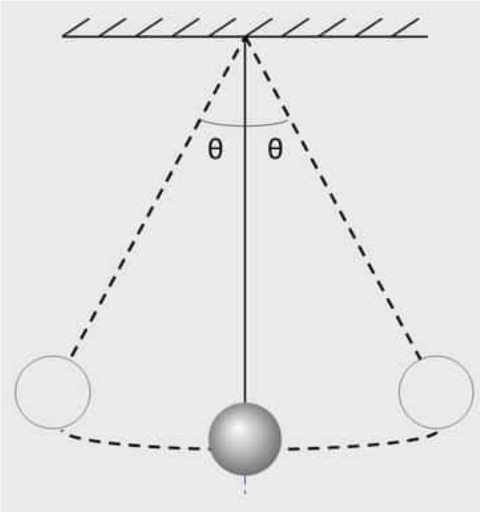
$$S_m^2 = \frac{\sum w_i}{\sum w_i \sum w_i x_i^2 - \left(\sum w_i x_i\right)^2}$$

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

$$S_b^2 = \frac{\sum w_i x_i^2}{\sum w_i \sum w_i x_i^2 - \left(\sum w_i x_i\right)^2}$$

Con $w_i = \frac{1}{S_{y_i}^2}$

Experimento: péndulo simple



Recordamos:

- Pequeñas oscilaciones
- hilo inextensible y de masa despreciable
- movimiento en el plano
- ¿Qué más?

El período del péndulo está dado por:

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

Actividad: Determinar la aceleración de la gravedad (g) a partir de los datos del periodo de un péndulo para distintas longitudes utilizando un modelo lineal

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

Hago un cambio de variables

T no está relacionada de forma lineal con la longitud l

Actividad: Determinar la aceleración de la gravedad (g) a partir de los datos del periodo de un péndulo para distintas longitudes utilizando un modelo lineal

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

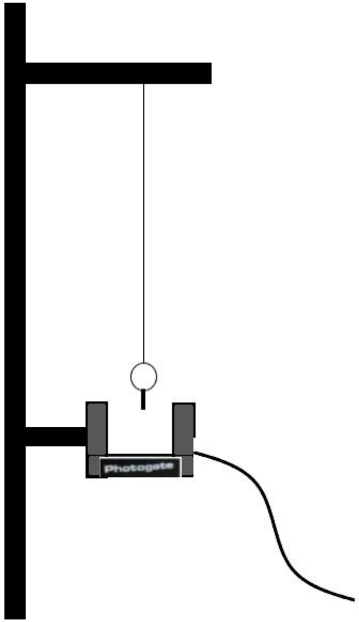
T no está relacionada de forma lineal con la longitud l

Hago un cambio de variables

$$\underset{y}{T} = \frac{2\pi}{\sqrt{g}}\underset{x}{\sqrt{l}} \longrightarrow m = \frac{2\pi}{\sqrt{g}}$$

$$\underset{y}{T^2} = \frac{4\pi^2}{g}\underset{x}{l} \longrightarrow m = \frac{4\pi^2}{g}$$

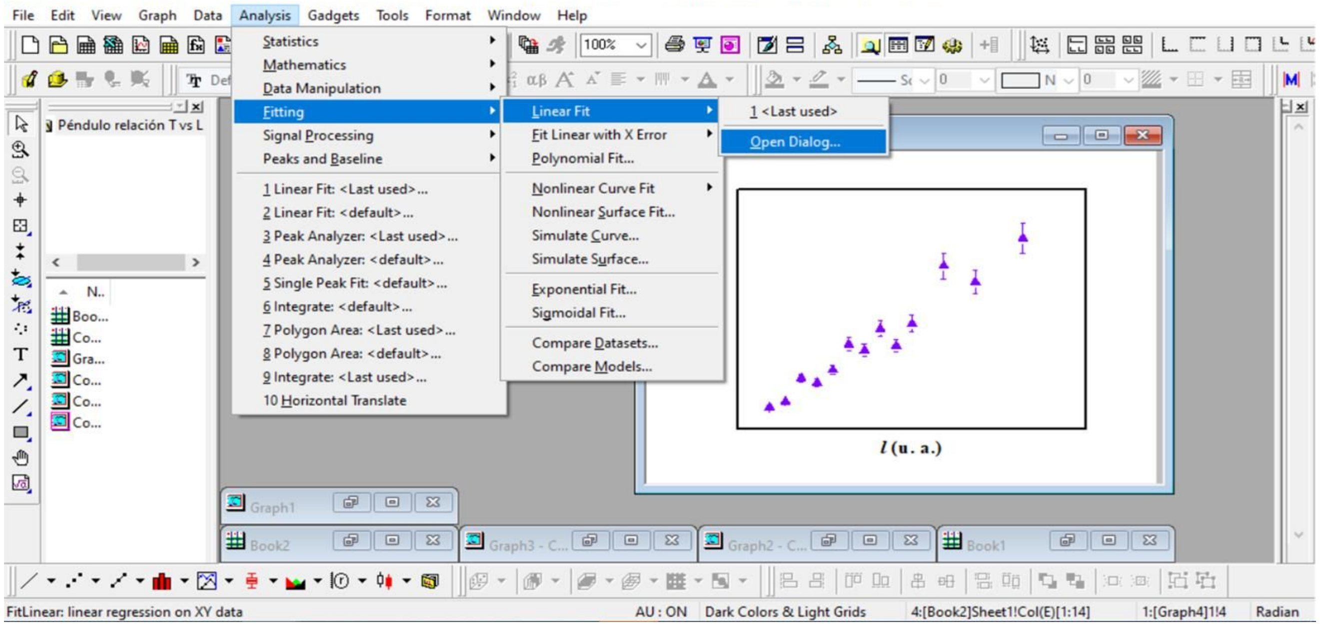
Actividad: Determinar la aceleración de la gravedad (g) a partir de los datos del periodo de un péndulo para distintas longitudes utilizando un modelo lineal



1. Preparar el experimento según el esquema
2. Tener en cuenta todas las suposiciones anteriores del péndulo simple.
3. Medir el periodo del péndulo.
4. Variar al menos 10 veces las longitudes del péndulo.
5. Calcular los errores absolutos de T^2 y l
6. Calcular los errores relativos de T^2 y l y compararlos. ¿Porqué?
7. Graficar T^2 vs l colocando las incertezas en la variable del eje y.
8. Realizar un ajuste lineal mediante el método de cuadrados mínimos (ponderado o no ponderados?)
9. Evaluar la calidad del ajuste (*coming soon*).
10. Obtener $\bar{g} \pm \Delta g$

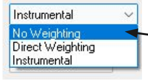
Se puede estudiar la precisión y la exactitud de su resultado?

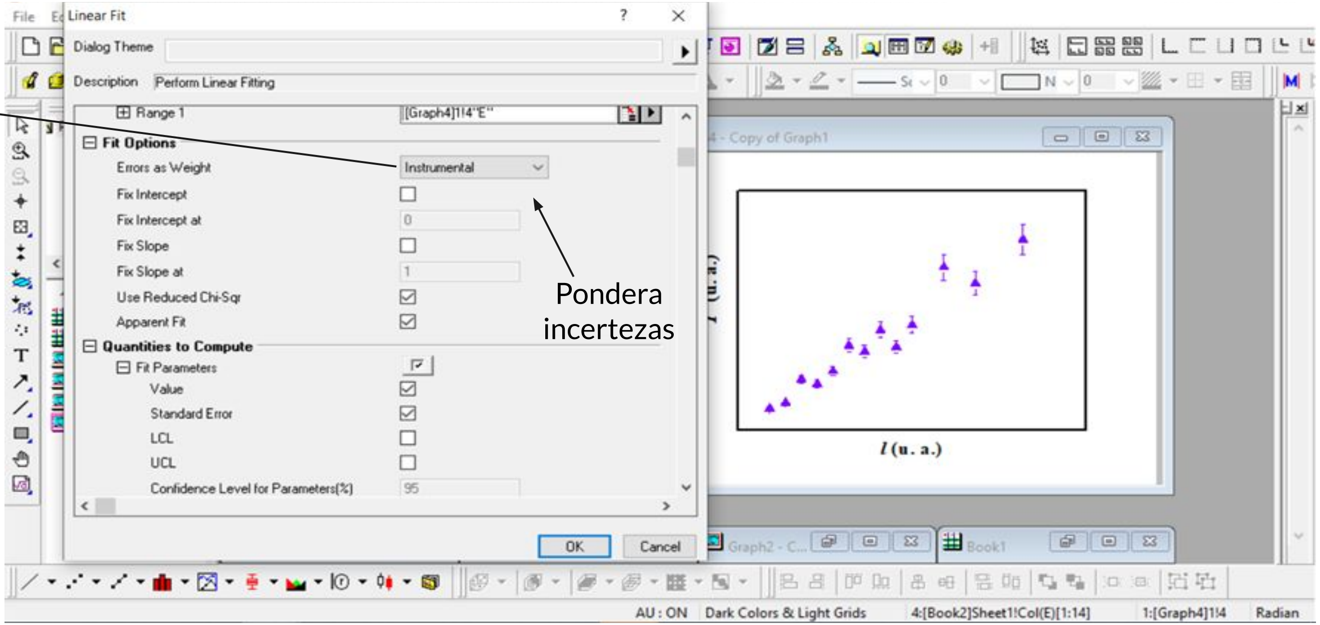
Ayuda para realizar un ajuste lineal en Origin



Método de cuadrados mínimos: Ponderados vs no ponderados

No ponderados





¿Cómo medimos la calidad del ajuste?

Coeficiente de Pearson r → Medida de la dependencia lineal entre dos variables

Chi-cuadrado χ^2 → Medida de la verosimilitud de un modelo o ajuste

Residuos → Distribución de los datos respecto de la recta

Coeficiente de correlación lineal de Pearson

Sean dos variables **x** e **y** de las cuales queremos ver la existencia de una correlación,

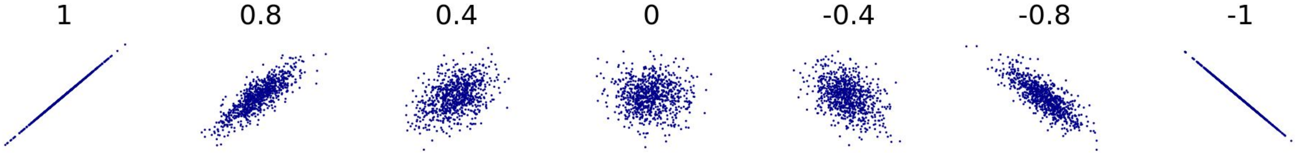
$$r = \frac{Cov(x,y)}{\sigma_x \sigma_y}; \quad \begin{cases} Cov(x,y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y}) \\ \sigma_x = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})}{N}} \\ \sigma_y = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \overline{y})}{N}} \end{cases}$$

En origin tenemos el R^2(adj Rsquare).

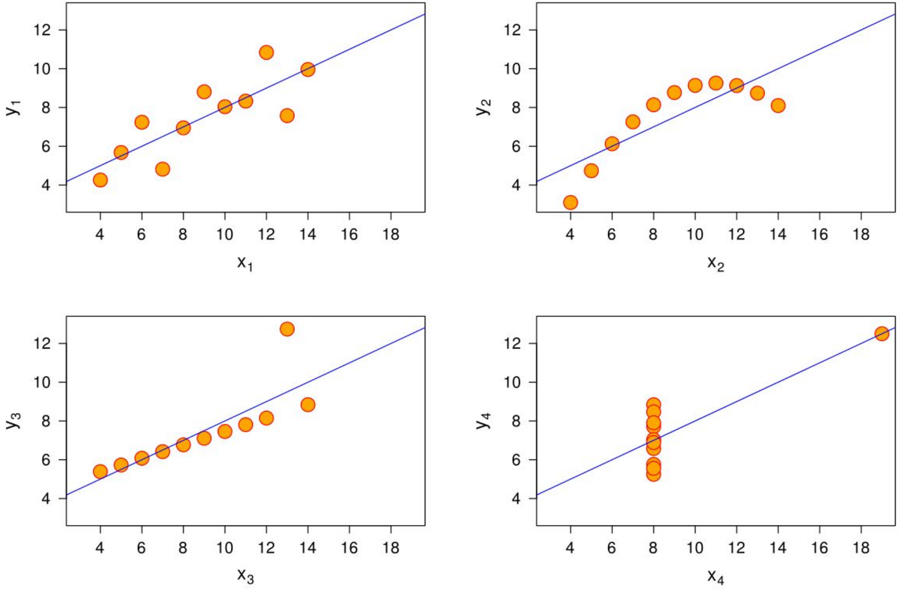
Si **x** e **y** están correlacionados linealmente:

$$|r| = 1$$

Su signo va a depender de la pendiente de la recta



Limitaciones del método: El cuarteto de Anscombe



Estos cuatro gráficos poseen las mismas propiedades estadísticas,

- Mismo valor medio y varianza de **x** y de **y**.
- Mismo R^2
- Estos datos ajustan exactamente igual de bien a una misma función lineal $y = 3 + 0{,}5x$

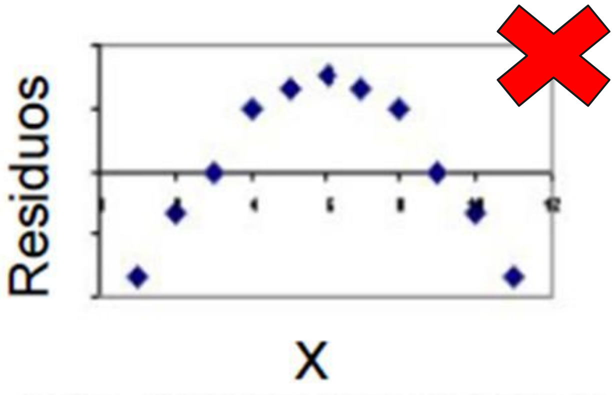
Propiedad	Valor
Media de cada una de las variables *x*	9.0
Varianza de cada una de las variables *x*	11.0
Media de cada una de las variables *y*	7.5
Varianza de cada una de las variables *y*	4.12
Correlación entre cada una de las variables *x* e *y*	0.816
Recta de regresión	$y = 3 + 0.5x$

Necesito usar un criterio extra!

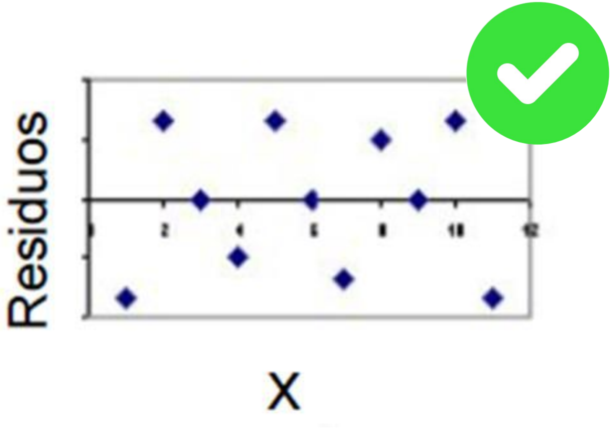
Adaptado de F.J. Anscombe, "Graphs in Statistical Analysis," American Statistician, 27 (1973), 17-21

Hay que mirar el gráfico: Residuos

Analizando el gráfico de los residuos en función de x:



Caso 2



Caso 1

En los casos 2,3 y 4, la distribución de los datos alrededor de la recta **no** era normal (gaussiana). Tenían estructura (no son aleatorios los puntos).

Chi-cuadrado χ^2

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_k)^2}{\sigma_i^2} \quad \cdots\cdots\rightarrow \quad \chi^2 = \sum_{i=1}^{N} \left[\frac{(y_i - (mx_i + b))}{S_i} \right]^2$$

→ Residuos

→ Medida del error

Con Oi los datos observados y Ek los datos ajustados. En general, para cualquier ajuste:

En un ajuste lineal, ajustamos m y b y comparamos contra N-2

$$\chi^2 \approx \underset{\uparrow}{\nu} = N - \underset{\uparrow}{k}$$

Grados de libertad

Cantidad de parámetros → Tenemos una relación entre las variables (vínculo)

Chi- cuadrado reducido

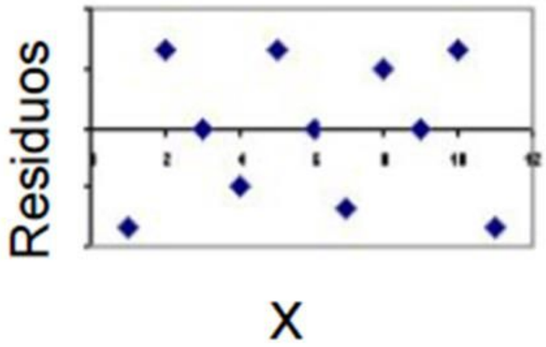
$$\chi^2_\nu = \frac{\chi^2}{\nu} = \frac{\chi^2}{N-k} \begin{cases} \chi^2_\nu \approx 1 \\ \chi^2_\nu << 1 \longrightarrow \text{Incertezas sobreestimadas} \\ \chi^2_\nu >> 1 \end{cases}$$

¿Qué utilizaremos?

$$|r| \approx 1$$

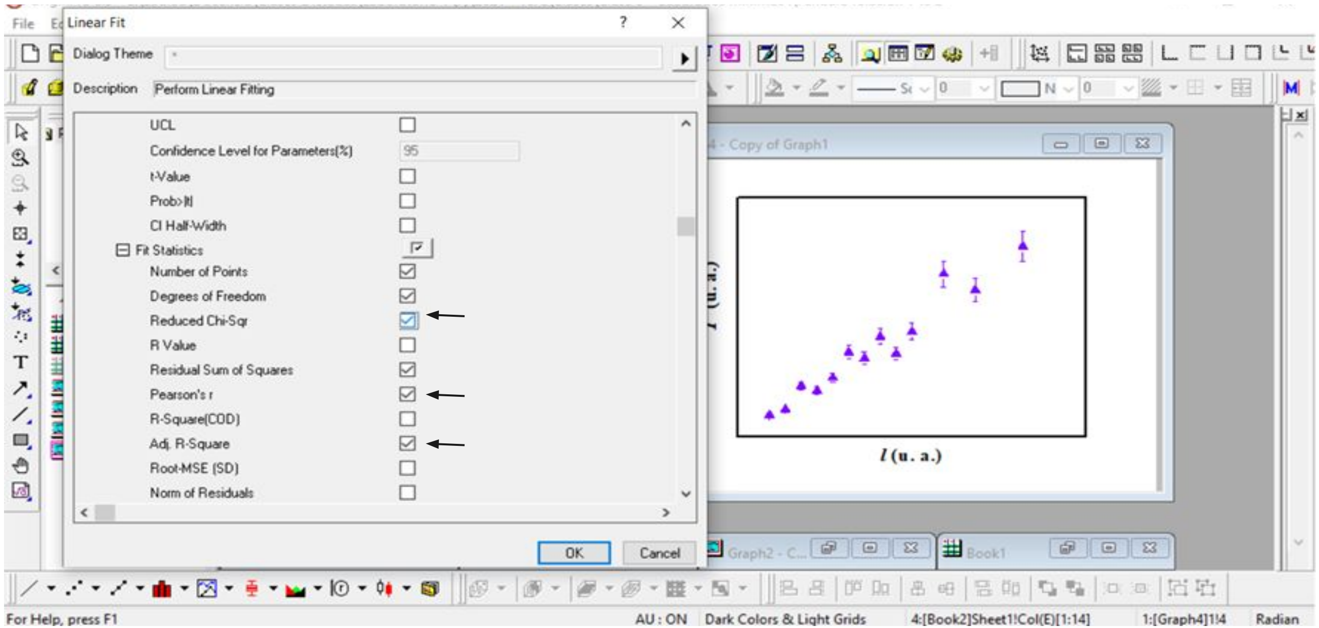
$$\chi^2_\nu \approx 1$$

Residuos

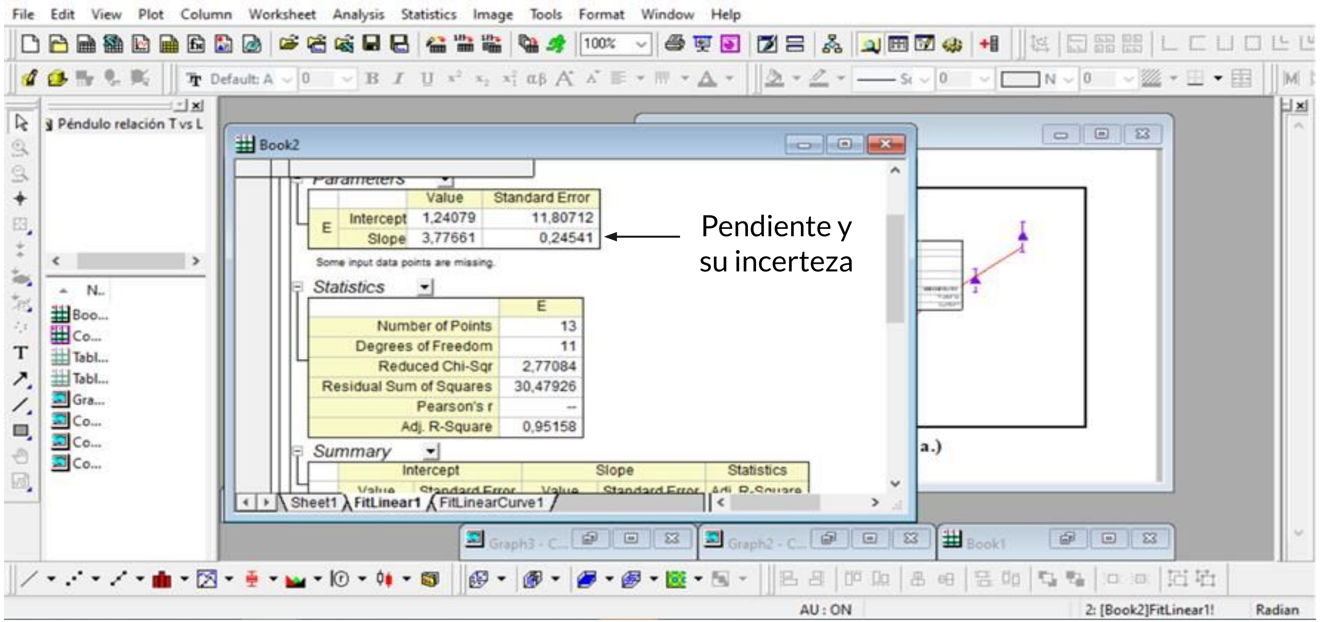


Estos métodos nos dan confianza en nuestro ajuste!

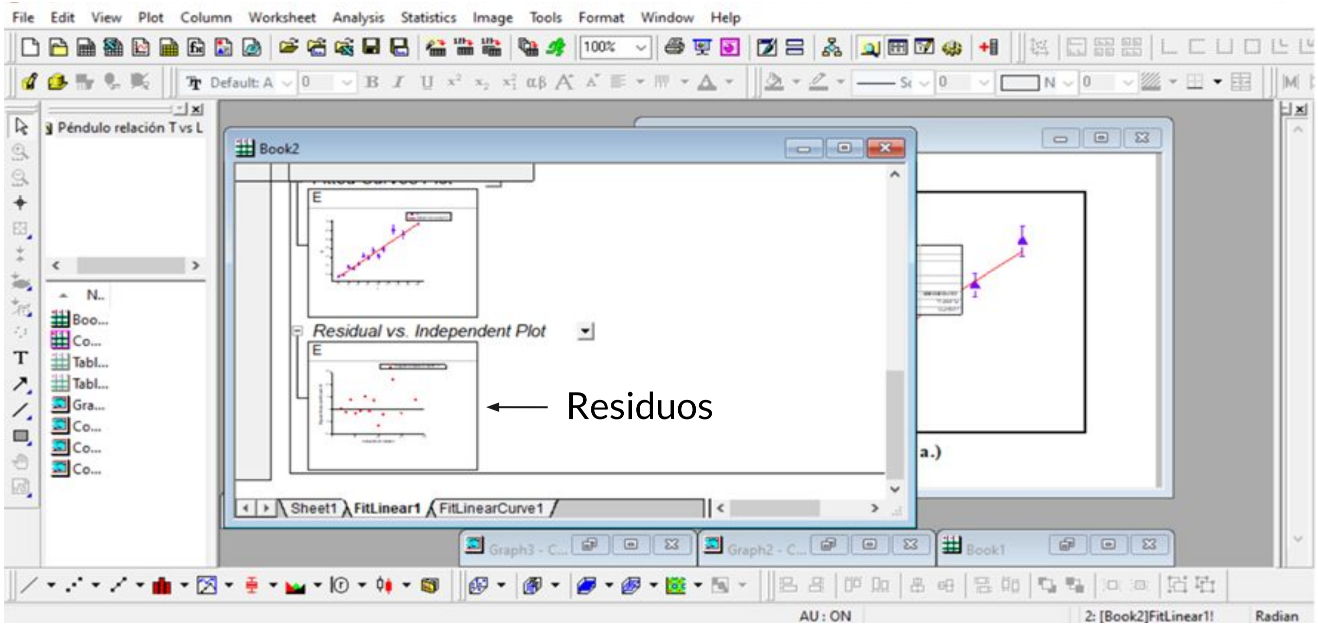
Calidad del ajuste en Origin



Resultados


File Edit View Plot Column Worksheet Analysis Statistics Image Tools Format Window Help
Péndulo relación T vs L
Book2
Value
Standard Error
Intercept 1,24079 11,80712
Slope 3,77661 0,24541
Some input data points are missing.
Statistics
Number of Points 13
Degrees of Freedom 11
Reduced Chi-Sqr 2,77084
Residual Sum of Squares 30,47926
Pearson's r --
Adj. R-Square 0,95158
Summary
Intercept
Slope
Statistics
Sheet1
FitLinear1
FitLinearCurve1
Pendiente y su incerteza
AU : ON
2: [Book2]FitLinear1!
Radian

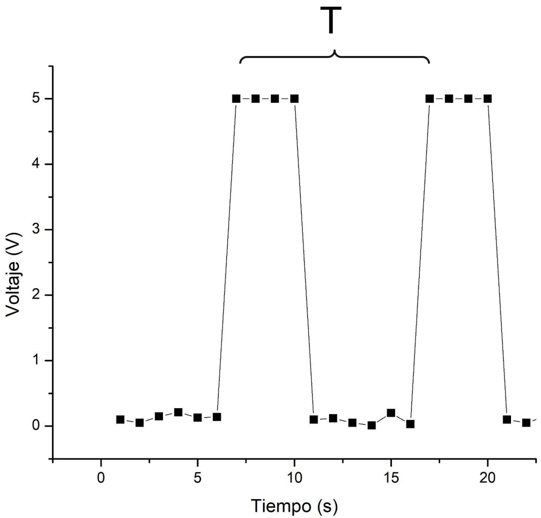
Residuos en Origin



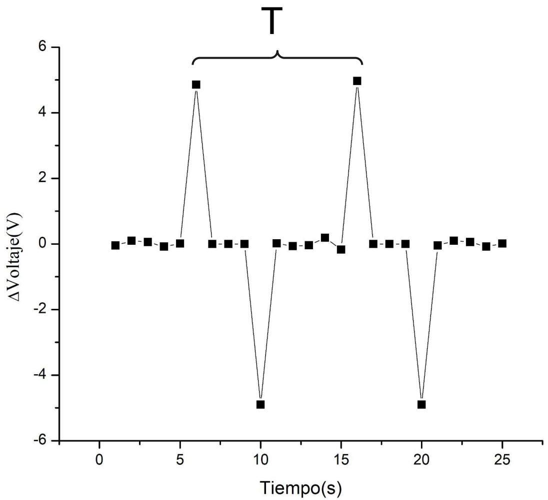
¿Preguntas?

Ayuda: Contemos cantidad de períodos desde el Origin

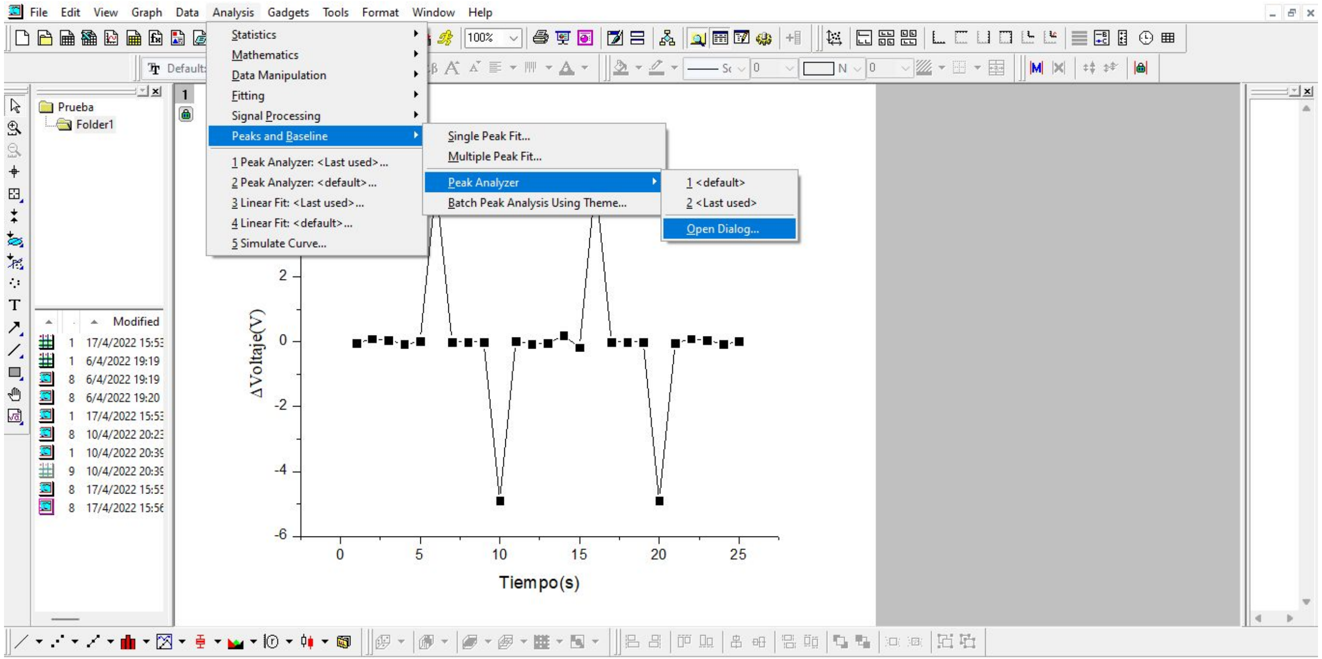
Ejemplo:



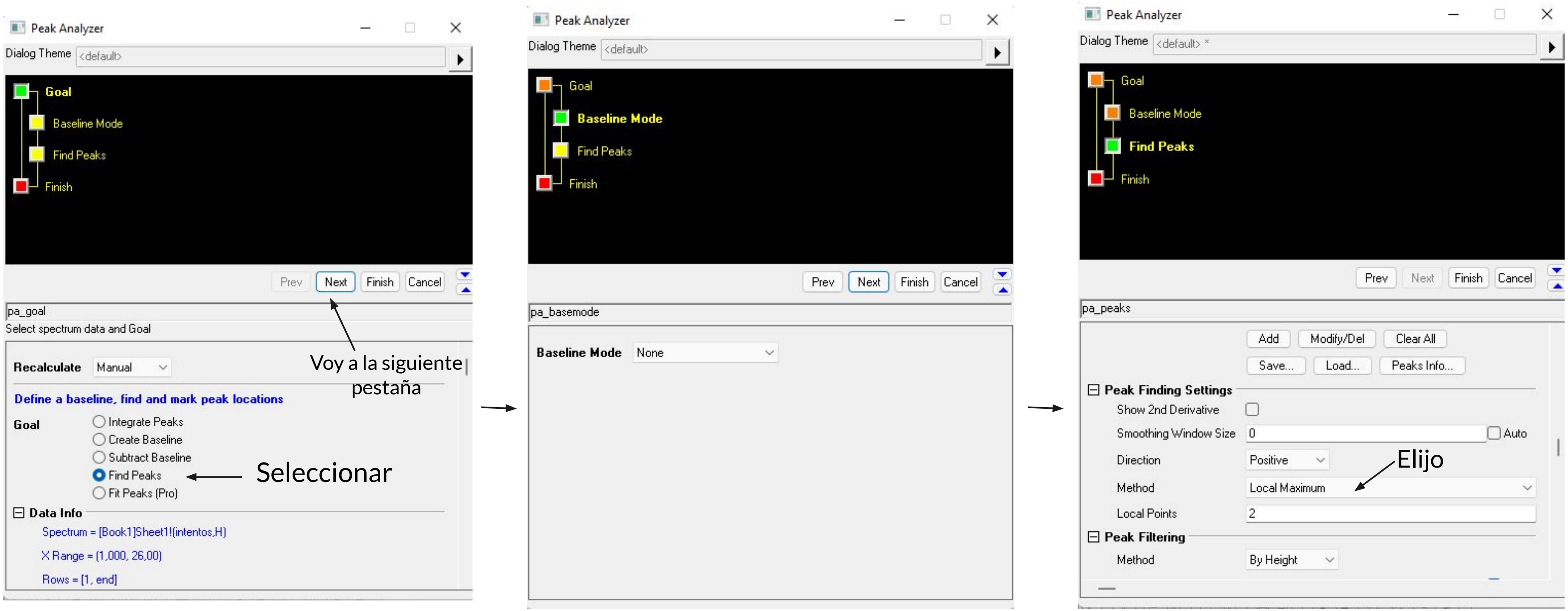
Set column Values → Col (B) [i+1]-Col(B)[i]



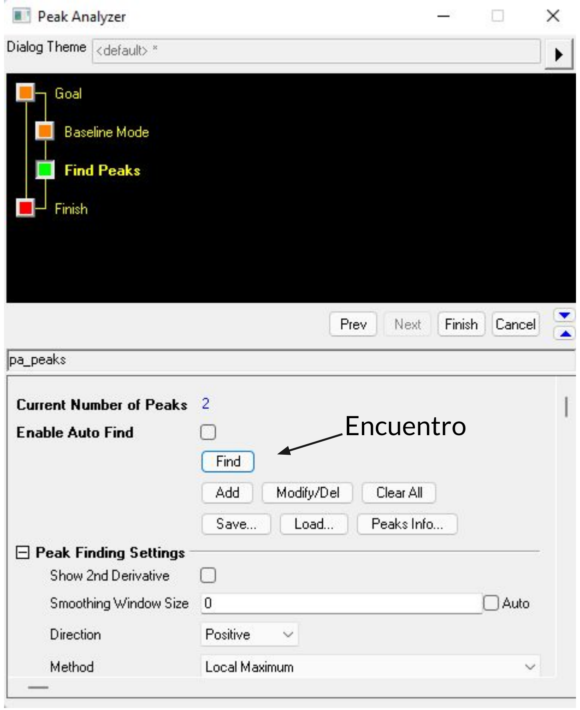
Encontramos los picos (máximos o mínimos):



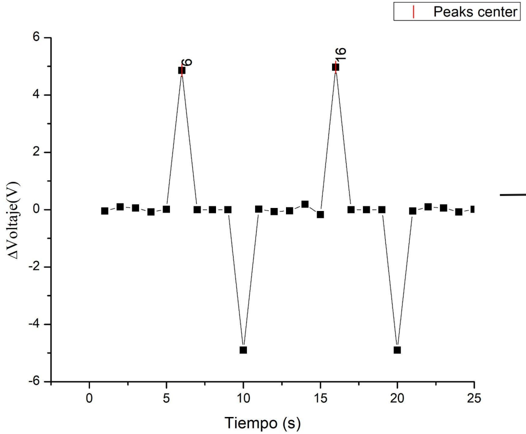
Pasos para encontrar los picos:



Finalmente,



Encuentro



Peak Cent X	Peak Cent Y
6	4,86
16	4,97

Contando la cantidad de periodos N:

$$T = \frac{T'}{N}$$