Modelo lineal y método de cuadrados mínimos

Laboratorio 1B – 2do Cuatrimestre 2024

Federico Trupp

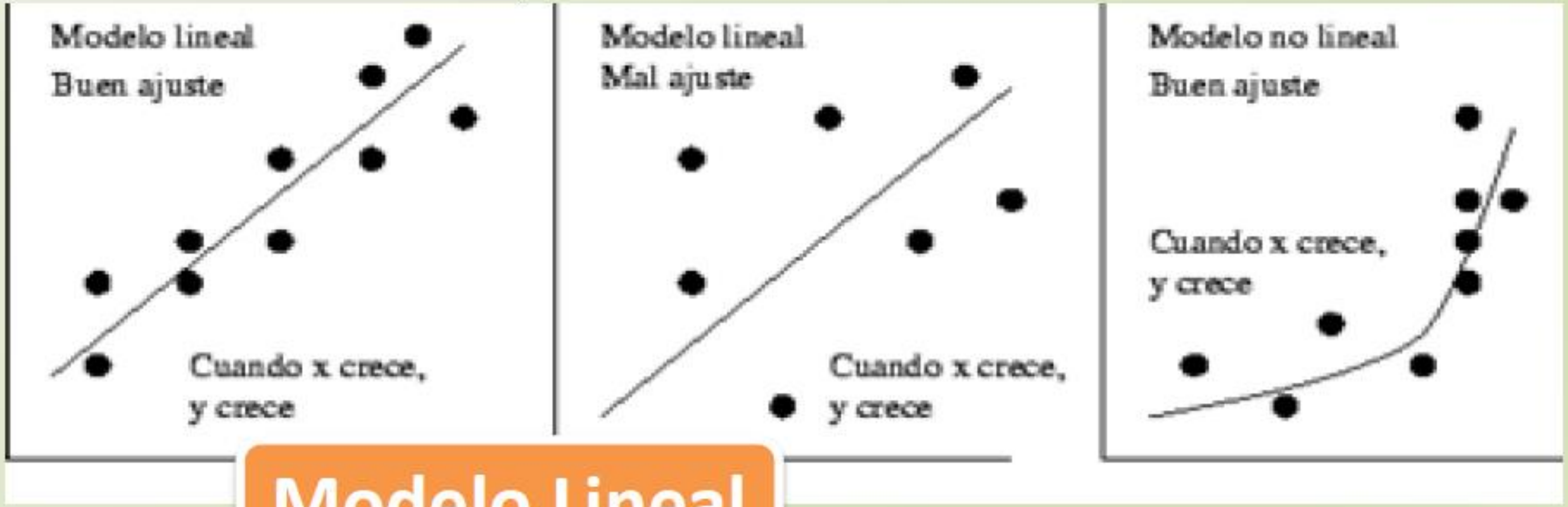
(basado en presentaciones Lucía Famá)

Modelado por el Método de Cuadrados Mínimos

Relación entre dos medidas:

Empleo un modelo que mejor relacione las variables (que mejor aproxime a mis datos experimentales)

Caso más sencillo



Modelo Lineal

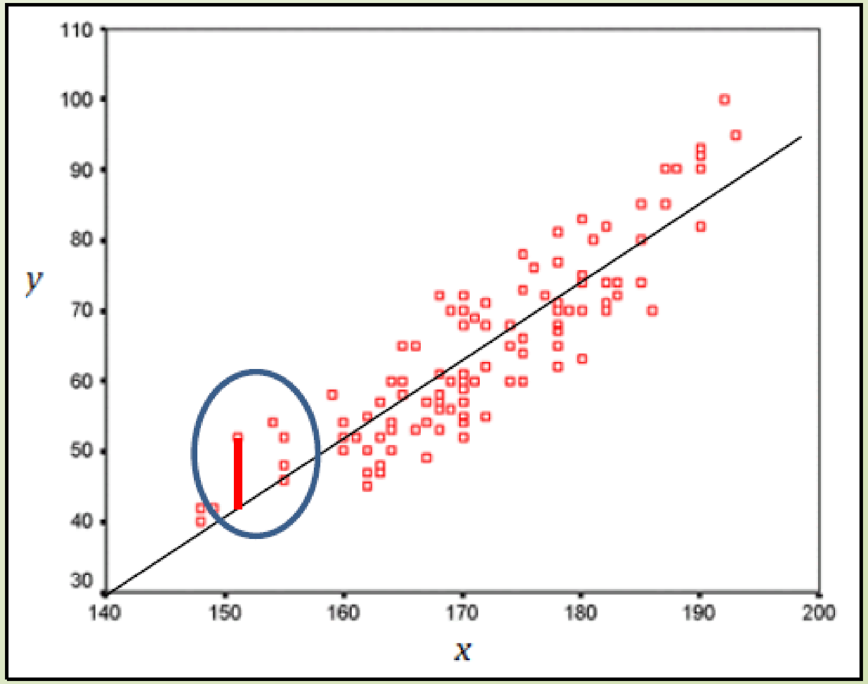
¿Por qué modelar?

- Se puede observar la relación entre las variables.
- El resultado es más representativo del sistema que tomar un único caso.
- Reducimos posibles fuentes de errores.
- Permite verificar la teoría.
- Se pueden realizar predicciones para la magnitud física fuera del rango de medición.

Caso más sencillo

Modelo Lineal

Tomamos una serie de medidas (x_i, y_i)



Partamos asumiendo que la relación es:

$$y = mx + b$$

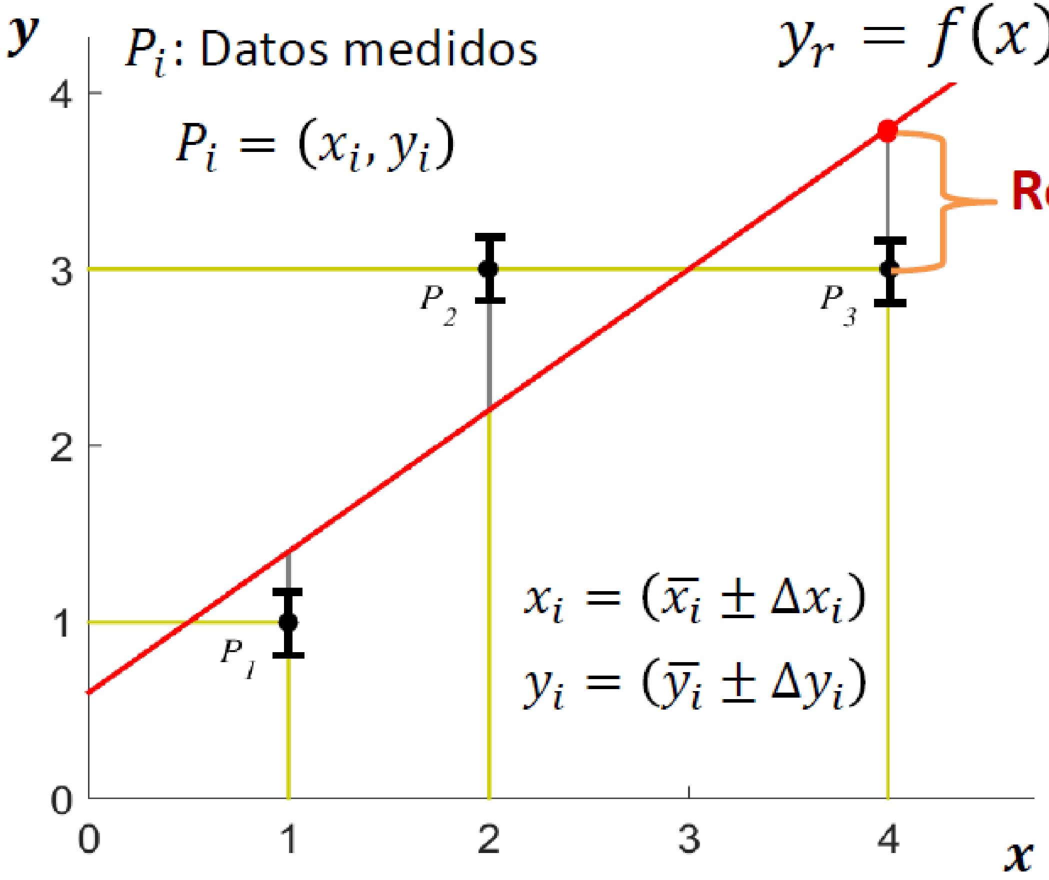
Caso aún más sencillo: Considerando la Distancia en "y"

Buscamos encontrar los parámetros m y b que minimicen la distancia de los datos al modelo en el eje "y"

Buscamos encontrar los parámetros m y b que minimicen la distancia de los datos al modelo

Considerando la Distancia en "y"

$$y_r = f(x) = mx + b$$



Residuos (δ_i) ⇨ $\delta_i = y_i - y_r$

Experimental **Recta**

Hipótesis:

$$\varepsilon_{ry} \gg \varepsilon_{rx}$$

$$\varepsilon_{ry} = \left|\frac{\Delta y}{y_o}\right|$$

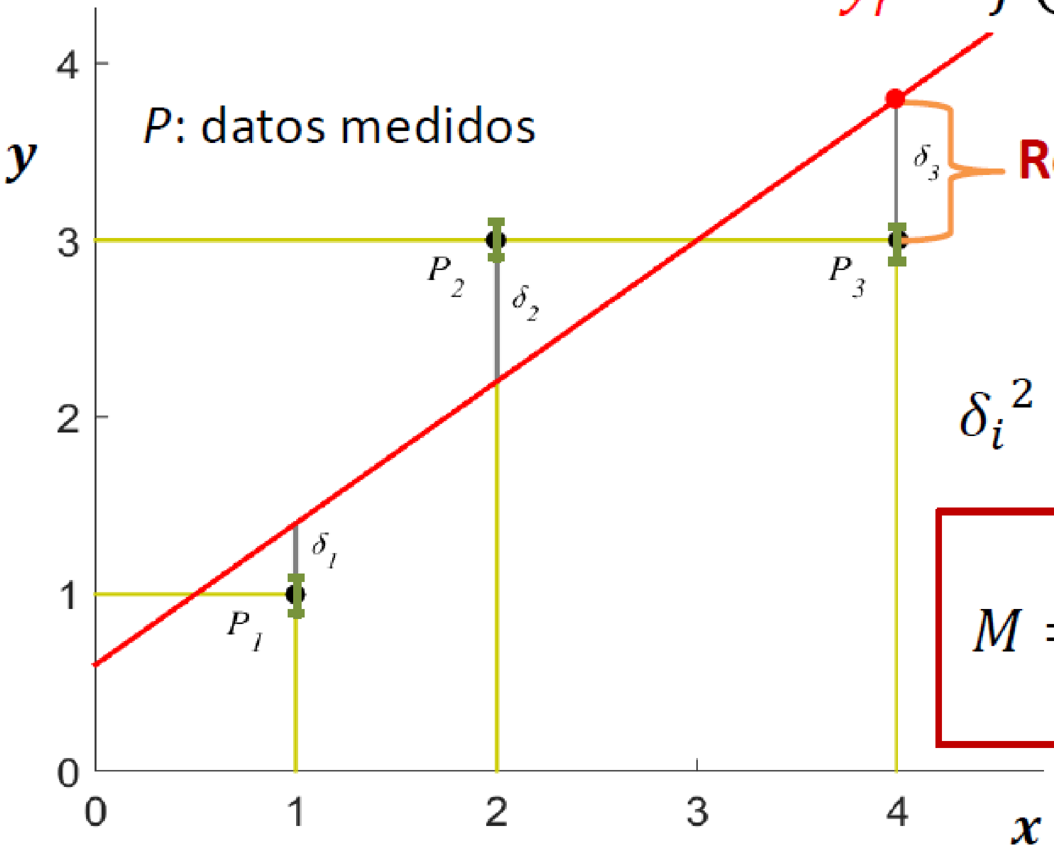
D. C. Baird. Prentice Hall (1991). Apéndice 2

Caso A

Cuadrados mínimos NO Ponderados

Cuando todos los datos en "*y*" tienen igual incerteza

$y_r = f(x) = mx + b$



Residuos (δ_i) ⇨ $\delta_i = y_i - y_r$

Experimental **Recta**

$$\delta_i{}^2 = [y_i - y_r]^2 = [y_i - (mx_i + b)]^2$$

$$\boxed{M = \sum_{i=1}^{N} \delta_i{}^2 = \sum_{i=1}^{N} [y_i - (mx_i + b)]^2}$$

Minimizar la Suma de los Cuadrados de los Residuos

¿Cómo encontramos los parámetros m y b?

Minimizar la Suma de los Cuadrados de los Residuos

$$M(m,b) = \sum_i y_i^2 + m^2 \sum_i x_i^2 + Nb^2 + 2mb \sum_i x_i - 2m \sum_i x_i y_i - 2b \sum_i y_i$$

$$\begin{cases} \dfrac{\partial M}{\partial m} = 0 \longrightarrow 2m \sum_i x_i^2 + 2b \sum_i x_i - 2 \sum_i x_i y_i = 0 \\ \dfrac{\partial M}{\partial b} = 0 \longrightarrow 2Nb + 2m \sum_i x_i - 2 \sum_i y_i = 0 \end{cases}$$

2 ecuaciones y
2 incógnitas

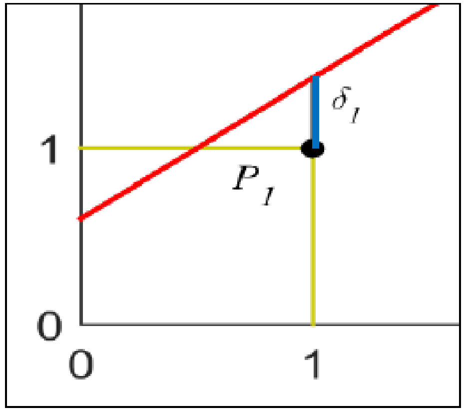
$$\boxed{m = \frac{N\sum x_i y_i - \sum x_i \sum y_i}{N\sum {x_i}^2 - (\sum x_i)^2}}$$

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

¿Cómo encontramos S_m y S_b?

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

Propagación de errores!! ⇨ *D. C. Baird. Prentice Hall (1991). Apéndice 2*



Estamos evaluando la incerteza en el eje y

→ *Hipótesis:* Consideremos a la incerteza como δ_i

$$S_m = S_y \sqrt{\frac{N}{N\sum {x_i}^2 - (\sum x_i)^2}} \qquad S_b = S_y \sqrt{\frac{\sum x_i}{N\sum {x_i}^2 - (\sum x_i)^2}} \quad \cdots\to \quad S_y = \sqrt{\frac{\sum {\delta_i}^2}{N-2}}$$

⇨ Calculamos las incertezas de forma exacta

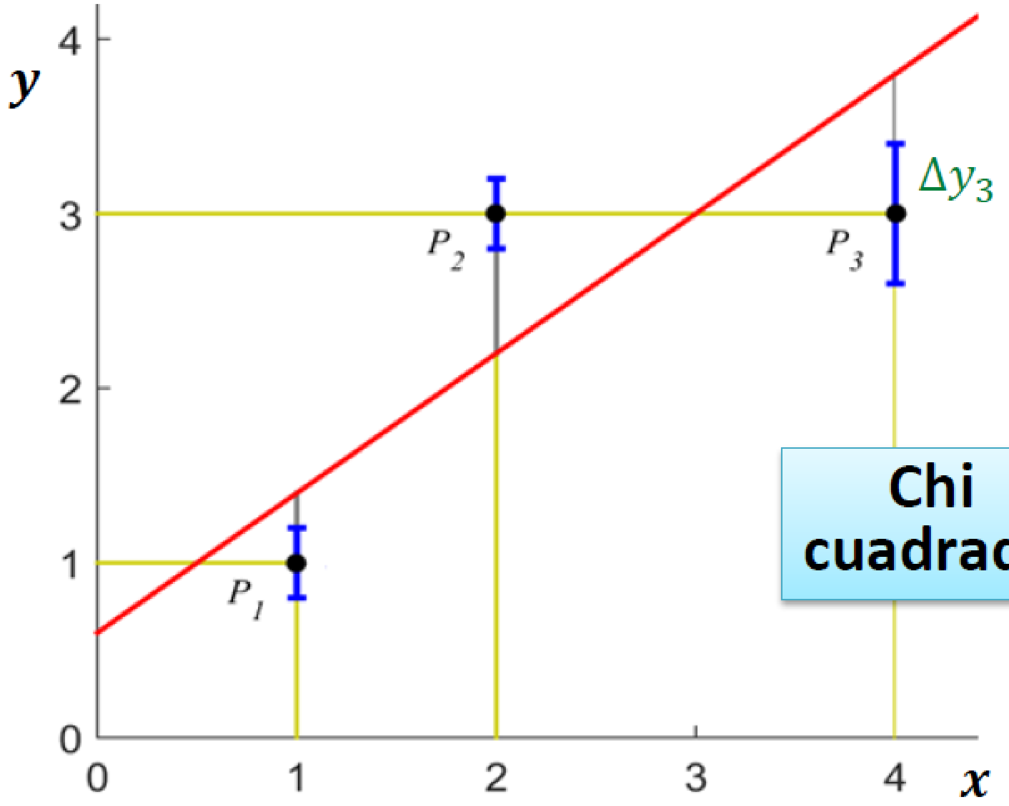
Válido cuando todas las medidas de y tienen igual precisión

Caso B

Cuadrados mínimos **Ponderados**

Cuando todos los datos en "y" tienen diferente incerteza

$$y = f(x) = mx + b$$



Hipótesis: Considera a las medidas más precisas como las más relevantes

Chi cuadrado

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

Minimizar la Suma de los Cuadrados de los Residuos
Normalizados al error de cada medida

Si confiamos más en algunos puntos que en otros

Cuadrados mínimos ponderados

Obtengo m y b despejando (2 ecuaciones y 2 incógnitas):

$$b = \frac{\sum \frac{1}{(S_{yi})^2} y_i \sum \frac{1}{(S_{yi})^2} {x_i}^2 - \sum \frac{1}{(S_{yi})^2} x_i \sum \frac{1}{(S_{yi})^2} x_i y_i}{\sum \frac{1}{(S_{yi})^2} \sum \frac{1}{(S_{yi})^2} {x_i}^2 - \left(\sum \frac{1}{(S_{yi})^2} x_i\right)^2}$$

$$m = \frac{\sum \frac{1}{(S_{yi})^2} \sum \frac{1}{(S_{yi})^2} (x_i y_i) - \sum \frac{1}{(S_{yi})^2} x_i \sum \frac{1}{(S_{yi})^2} y_i}{\sum \frac{1}{(S_{yi})^2} \sum \frac{1}{(S_{yi})^2} {x_i}^2 - \left(\sum \frac{1}{(S_{yi})^2} x_i\right)^2}$$

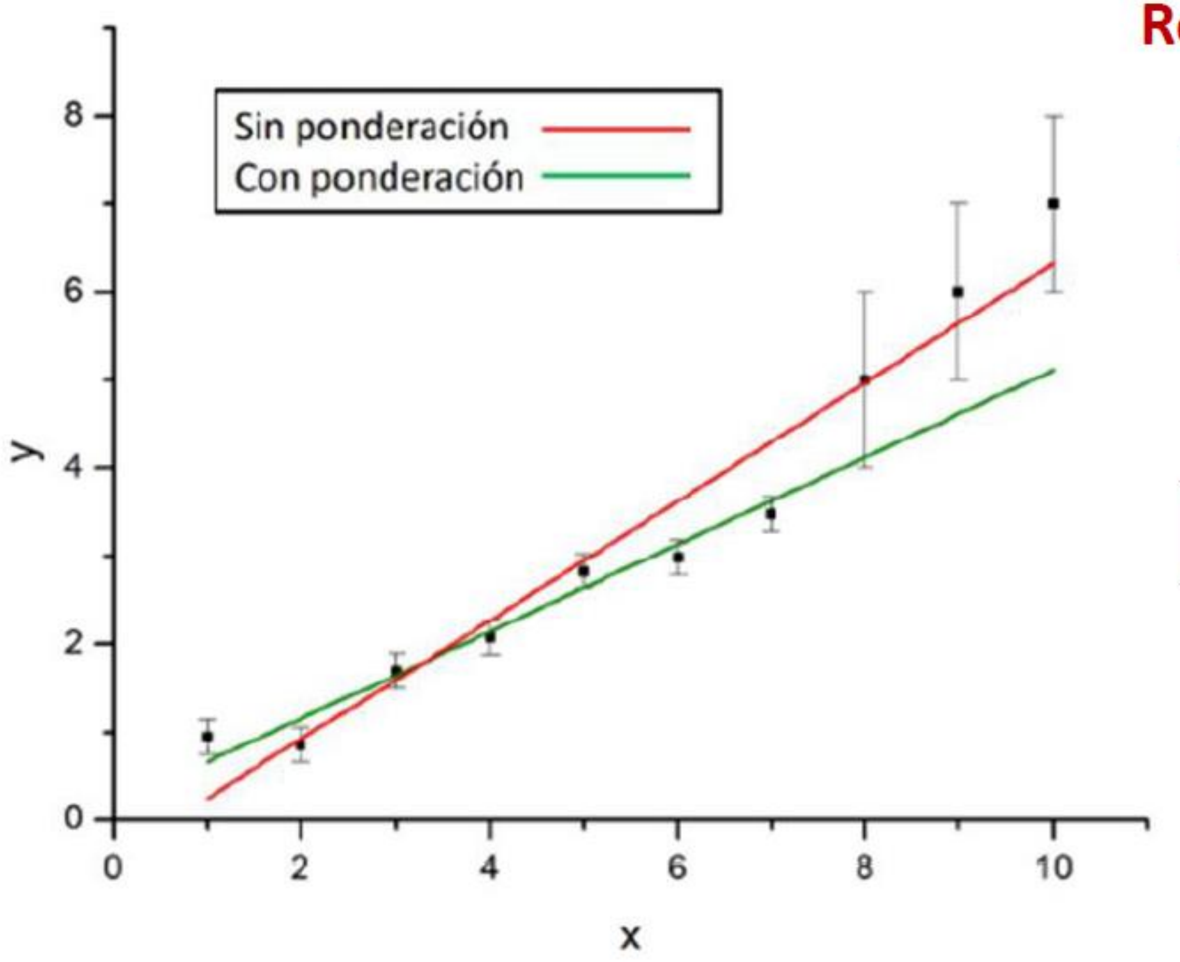
$${S_m}^2 = \frac{\sum \frac{1}{(S_{yi})^2}}{\sum \frac{1}{(S_{yi})^2} \sum \frac{{x_i}^2}{(S_{yi})^2} - \left(\sum \frac{x_i}{(S_{yi})^2}\right)^2}$$

$${S_b}^2 = \frac{\sum \frac{x_i y_i}{(S_{yi})^2}}{\sum \frac{1}{(S_{yi})^2} \sum \frac{{x_i}^2}{(S_{yi})^2} - \left(\sum \frac{x_i}{(S_{yi})^2}\right)^2}$$

Lo importante es que puedo calcular estos valores de manera **EXACTA**

SIN Ponderación vs CON Ponderación

Se pondera cuando las incertezas absolutas de la MF que se asigna al eje "y" tienen diferente precisión



Residuos (δ_i) ➡ $\delta_i = y_i - y_r$

SIN

$$M = \sum_{1}^{N} [y_i - (mx_i + b)]^2$$

CON

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

Al ponderar, considera más relevantes a las medidas más precisas

¿Cómo determinamos si el ajuste es “bueno”?
¿Qué medidas podemos evaluar y comparar?

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

¿Existe algún patrón entre ellas?

$$r = \frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}} = \frac{S_{xy}}{S_x S_y}$$

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

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

$$-1 \leq r \leq 1$$

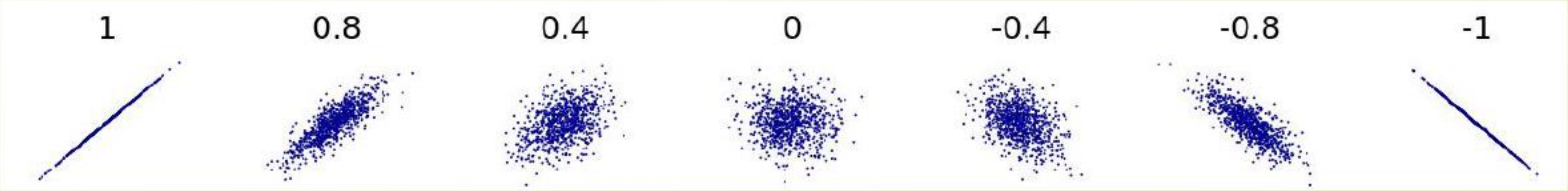
$$Cov(x,y) = S_{xy} = \frac{1}{N}\sum (x_i - \bar{x})(y_i - \bar{y})$$

Se espera que $|r| \sim 1$

Cálculo del Correlación de Pearson en Python: **corrcoef()**

Parámetros de BONDAD

Coeficiente de Correlación de Pearson (r)



- **Si r = 1:** Correlación positiva perfecta. El índice refleja la dependencia total entre ambas dos variables, la que se denomina relación directa: cuando una de las variables aumenta, la otra variable aumenta en proporción constante.

- **Si 0 < r < 1:** Refleja que se da una correlación positiva.

- **Si r = 0:** En este caso no hay una relación lineal. Aunque no significa que las variables sean independientes, puede haber relaciones no lineales entre ambas.

- **Si -1 < r < 0:** Indica que existe una correlación negativa.

- **Si r = -1:** Indica una **correlación negativa perfecta** y una dependencia total entre ambas variables lo que se conoce como **"relación inversa"**, que es cuando una de las variables aumenta, la otra variable disminuye en proporción constante.

Vamos a usar el coeficiente de determinación R^2

En el caso de los modelos lineales, es igual al coeficiente de Pearson al cuadrado

$$R^2 = \frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2} = \rho^2$$

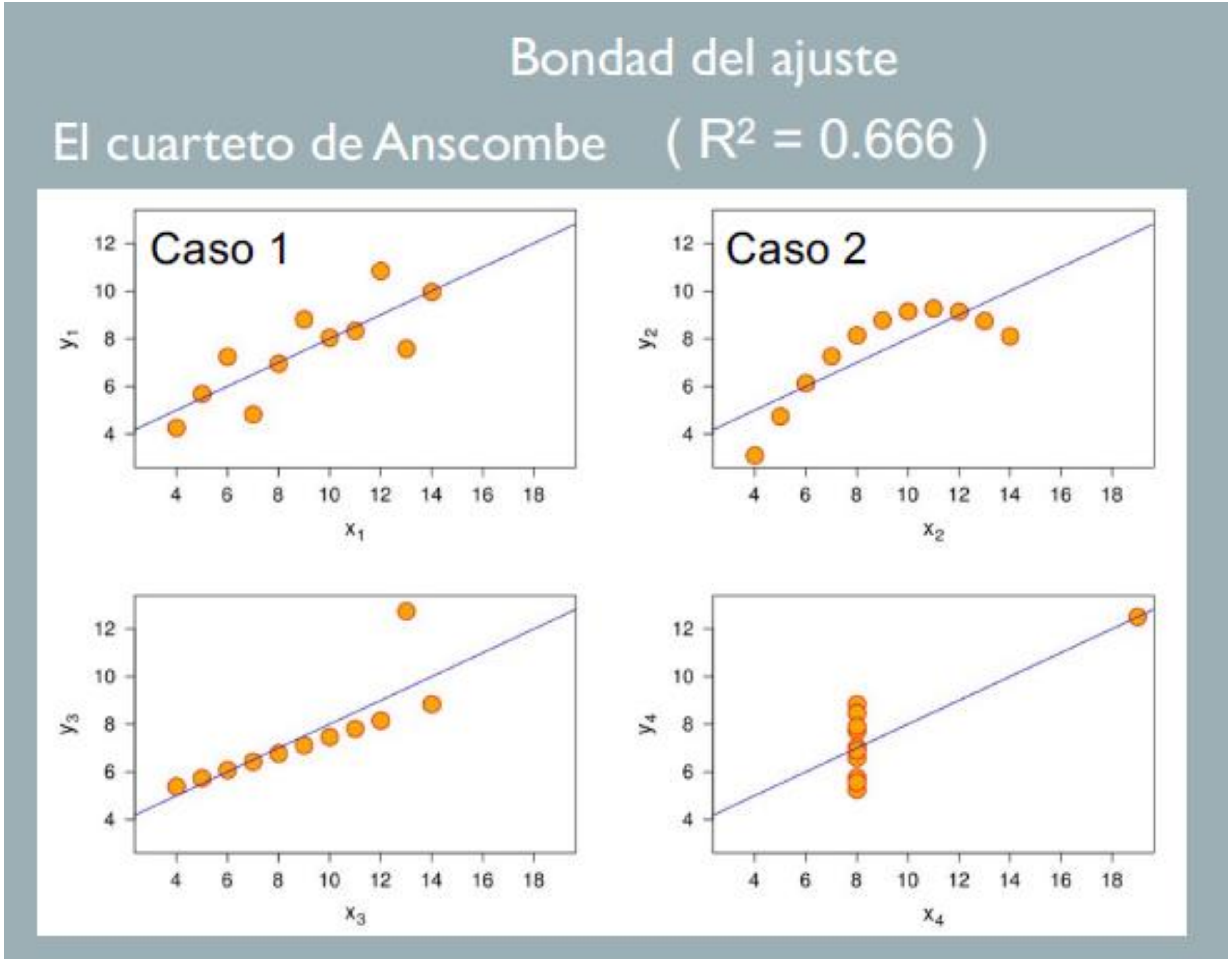
$$\rho^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} = 1 - \frac{\sigma_r^2}{\sigma_y^2}$$

Donde:

- σ_{XY} es la covarianza de (X, Y)
- σ_X^2 es la varianza de la variable X
- σ_Y^2 es la varianza de la variable Y

R^2 está entre 0 y 1
Cuanto más cerca de 1, mayor es la correlación

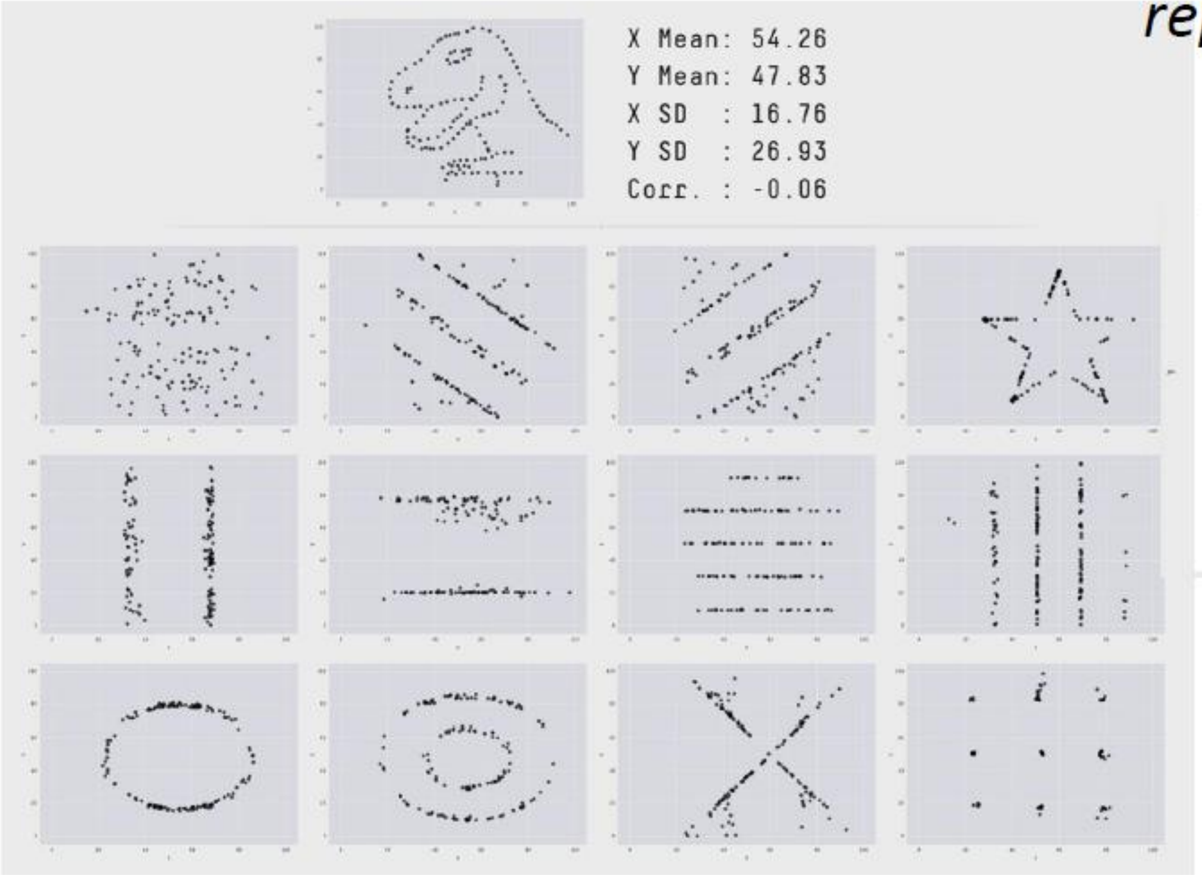
Pero ojo! el R^2 solo no alcanza para determinar la bondad de un ajuste

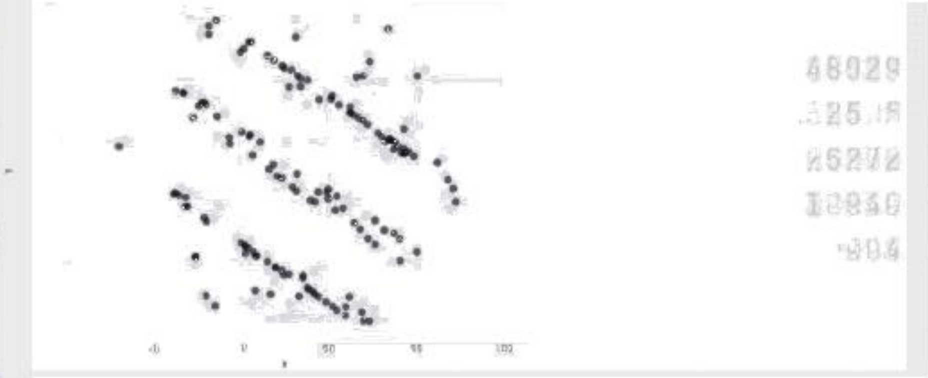


Estos casos tienen los MISMOS Parámetros Estadísticos!!!

Datasaurus dozen

... pero que producen gráficos diferentes
Por eso la IMPORTANCIA de las representaciones gráficas



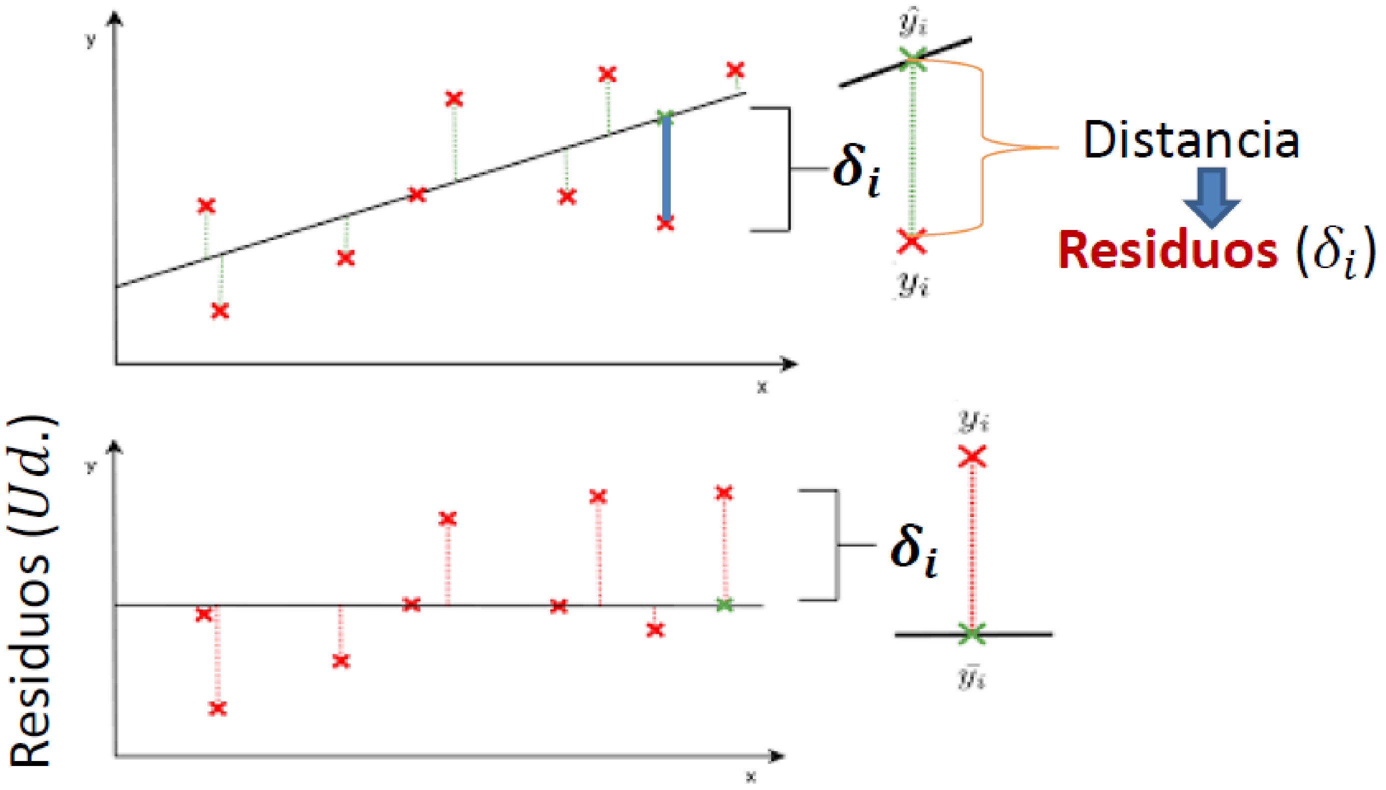


https://www.autodesk.com/research/publications/same-stats-different-graphs

Parámetros de BONDAD

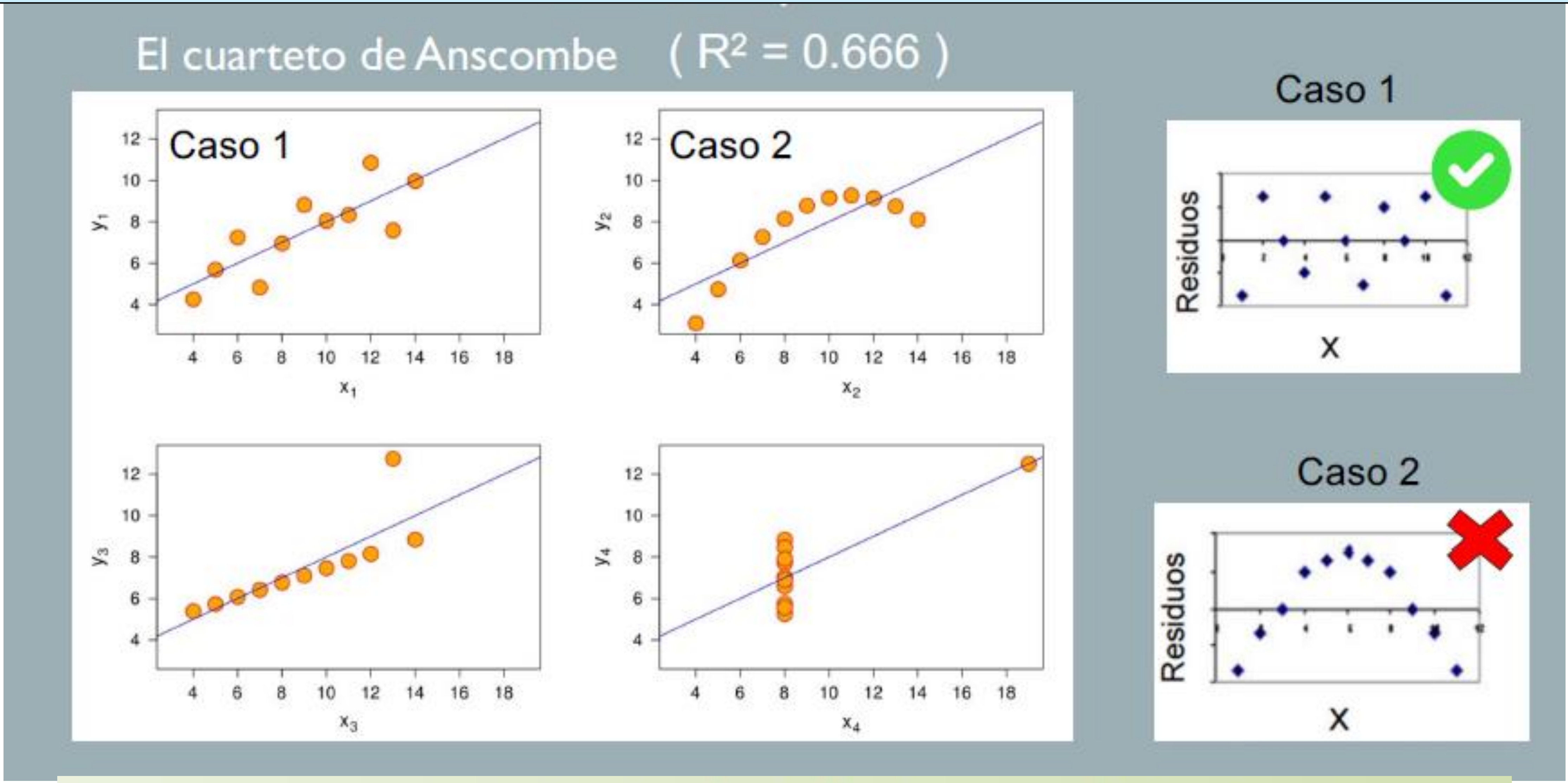
Gráfico de Residuos

Distancia de los puntos experimentales a la recta, en "y"



Los residuos deben ser datos distribuidos en forma aleatoria alrededor del cero.
NO DEBEN tener ESTRUCTURA

También tenemos que observar el gráfico de residuos y ver si tienen estructura



En los casos 2, 3 y 4 la distribución de los datos alrededor de la recta no es normal. **Los residuos tienen estructura**

Por último, necesitamos una medida que nos dé información sobre nuestra estimación de incertezas: χ_ν^2(chi cuadrado reducido)

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

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

Caso lineal:

N = número de datos

2: los grado de libertad

Chi cuadrado reducido ➡ $\chi_v^2 = \dfrac{\chi^2}{N-2}$

Se espera que χ_v^2

$\chi_v^2 \sim 1$ ✓

$\chi_v^2 \ll 1$ ✗

$\chi_v^2 \gg 1$ ✗

Parámetros de BONDAD

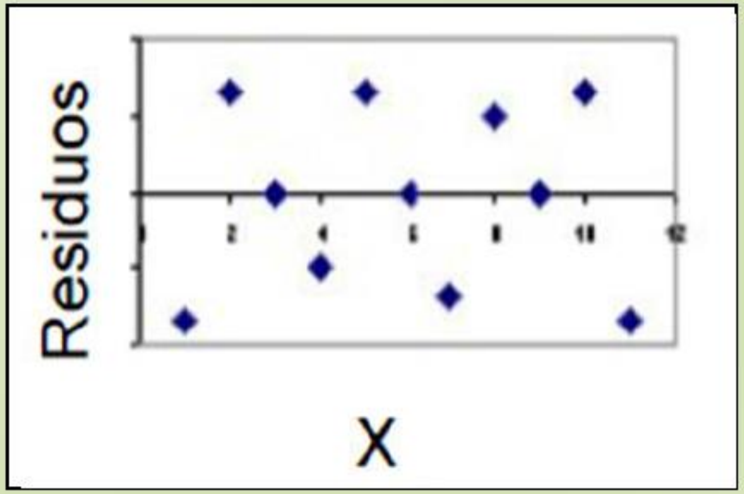
¿Qué esperamos?

Parámetros de BONDAD que nos servirán de ayuda:

Coeficiente de Determinación R^2 → $R^2 \sim 1$

Chi-cuadrado reducido χ^2_v → $\chi^2_v \sim 1$

Residuos Sin ESRUCTURA →



Objetivo de la práctica de hoy

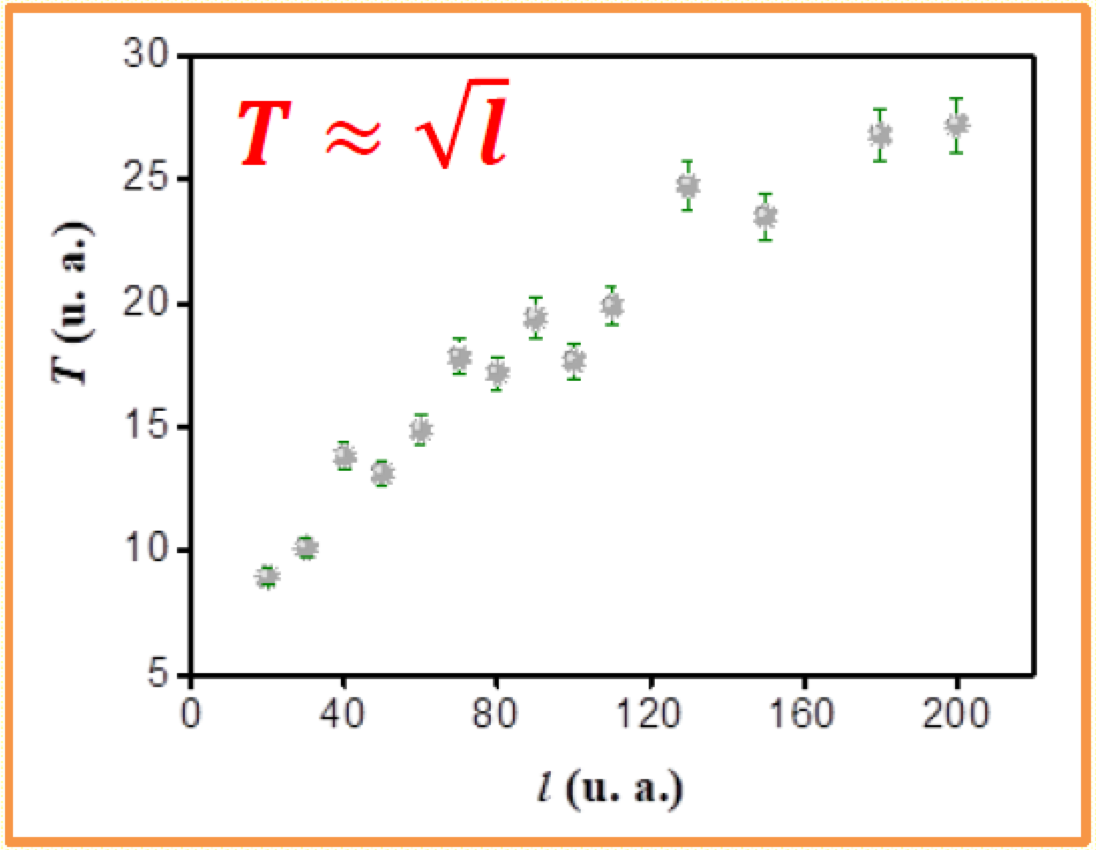
Determinar la aceleración de la gravedad g a partir de los datos del Período de un Péndulo T y su Longitudes l

$$\boxed{g = (\overline{g} \pm \Delta g)\ \boldsymbol{Ud.}}$$

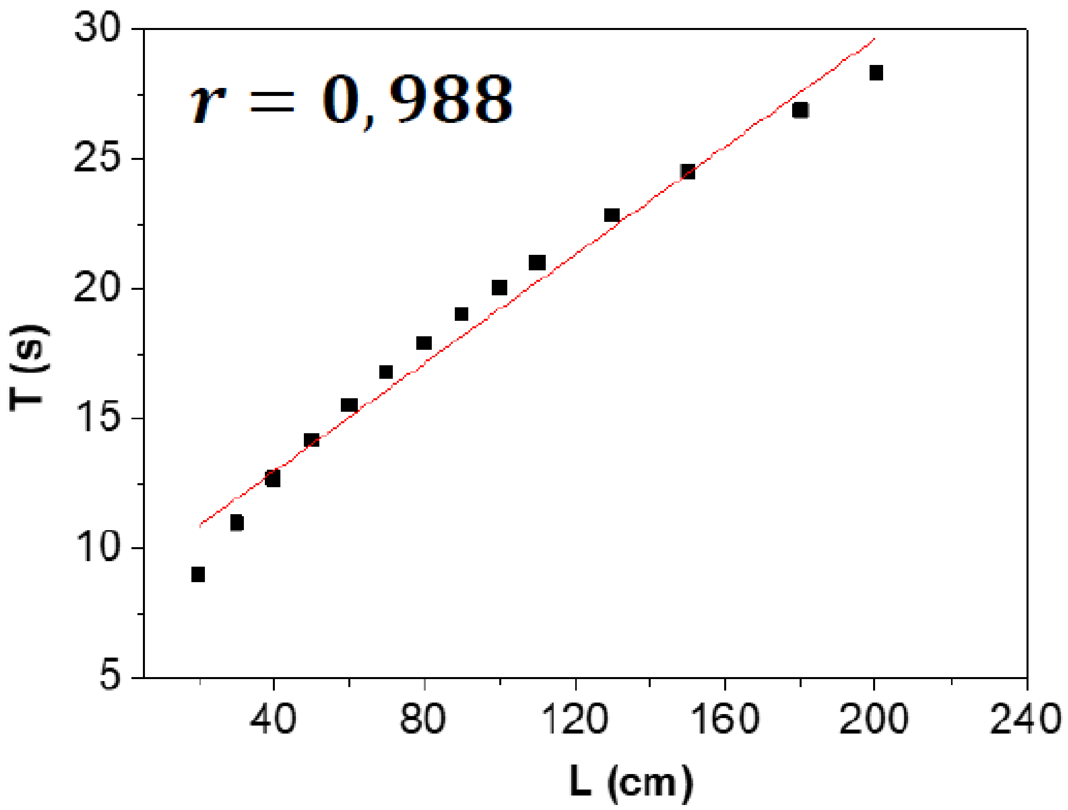
Determinar la aceleración de la gravedad g a partir de T y l de DIFERENTES Péndulos y un MODELO LINEAL DEL MÉTODO DE CUADRADOS MÍNIMOS

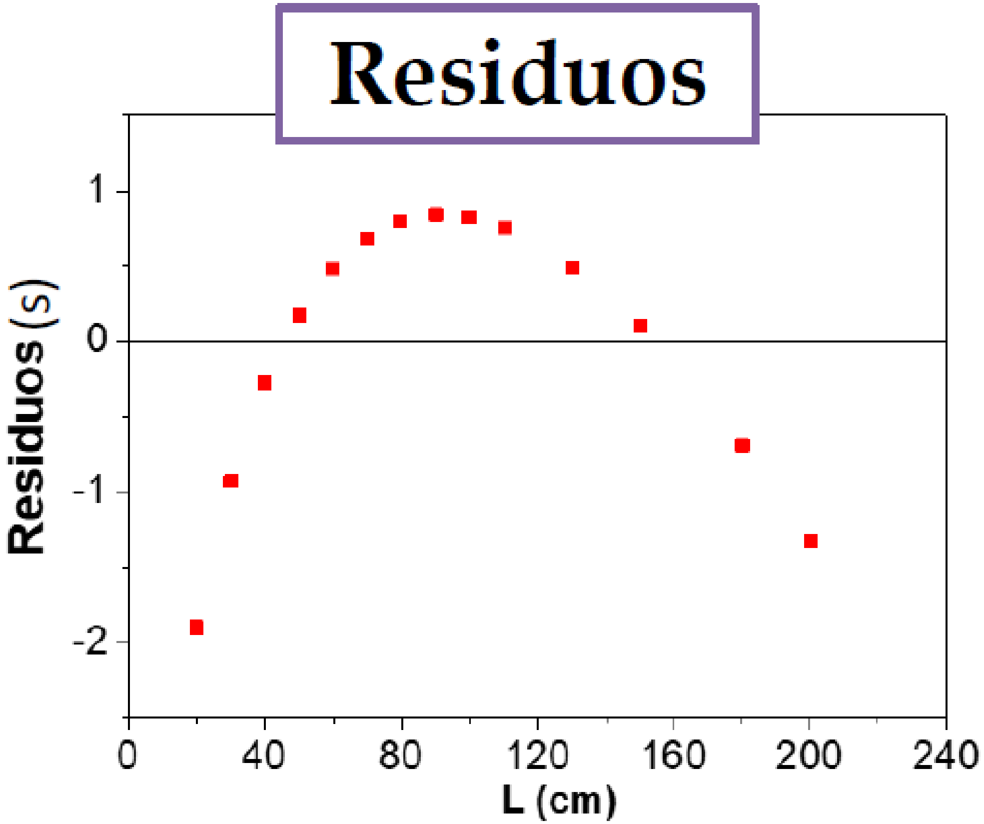
PERO T Y l SE RELACIONAN LINEALMENTE??

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



Está bien usar el modelo lineal en este caso?





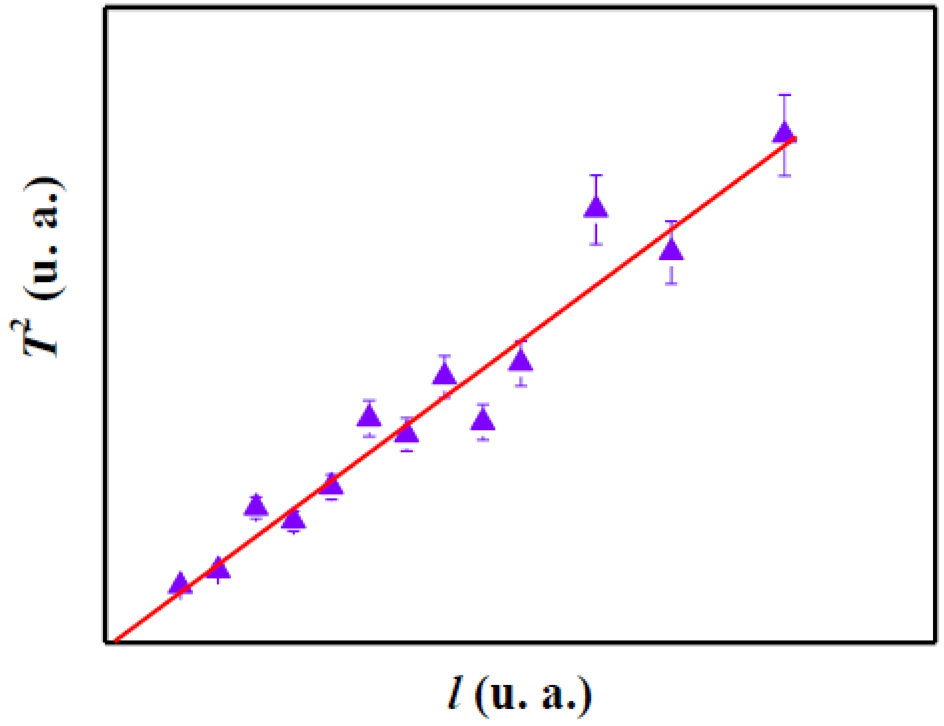
NOOOOOO es lineal!!

¿Cómo utilizo el modelo lineal en una relación NO lineal?

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

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

$$\underset{\tilde{T}}{\underset{\uparrow}{T^2}} = \underbrace{\frac{4\pi^2}{g}}_{\text{Pendiente}}\underset{l}{\underset{\uparrow}{l}}$$



✓ $y = mx + b$

Queremos obtener: $\boxed{g = (\bar{g} \pm \Delta g)\ Ud.}$

Uno de los dos casos:

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

y x

Pendiente $m = \bar{m} + \Delta m$

$$m = \frac{4\pi^2}{g} \longrightarrow g = \frac{4\pi^2}{m}$$

$$\boxed{\bar{g} = \frac{4\pi^2}{\bar{m}}}$$

¿ Δg?

Propago!!

$$\boxed{\Delta y = \sqrt{\left(\left.\frac{\partial f}{\partial p}\right|_{y_0} \Delta p\right)^2 + \left(\left.\frac{\partial f}{\partial q}\right|_{y_0} \Delta q\right)^2 + \left(\left.\frac{\partial f}{\partial r}\right|_{y_0} \Delta r\right)^2 + \cdots}}$$

Son productos y divisiones

$$y = \frac{p}{q} \qquad \frac{\Delta y}{y_0} = \sqrt{\frac{(\Delta q)^2}{{q_0}^2} + \frac{(\Delta p)^2}{{p_0}^2}}$$

$$y = p \cdot q \qquad \frac{\Delta y}{y_0} = \sqrt{\frac{(\Delta q)^2}{{q_0}^2} + \frac{(\Delta p)^2}{{p_0}^2}}$$

Para la próxima clase, se entrega el cuaderno con:

1. **PRÁCTICA 2 - ACTIVIDAD 2**: DETERMINACIÓN DE G A PARTIR DE LA MEDICIÓN DEL PERÍODO DE UN PÉNDULO (con los datos del TP1)

2. **PRÁCTICA 3 - ACTIVIDAD 1**: DETERMINACIÓN DE G A PARTIR DE LA MEDICIÓN DEL PERÍODO DE UN PÉNDULO (actividad de hoy)

Como siempre: antes de las 13 hs del miércoles que viene en el mismo link.