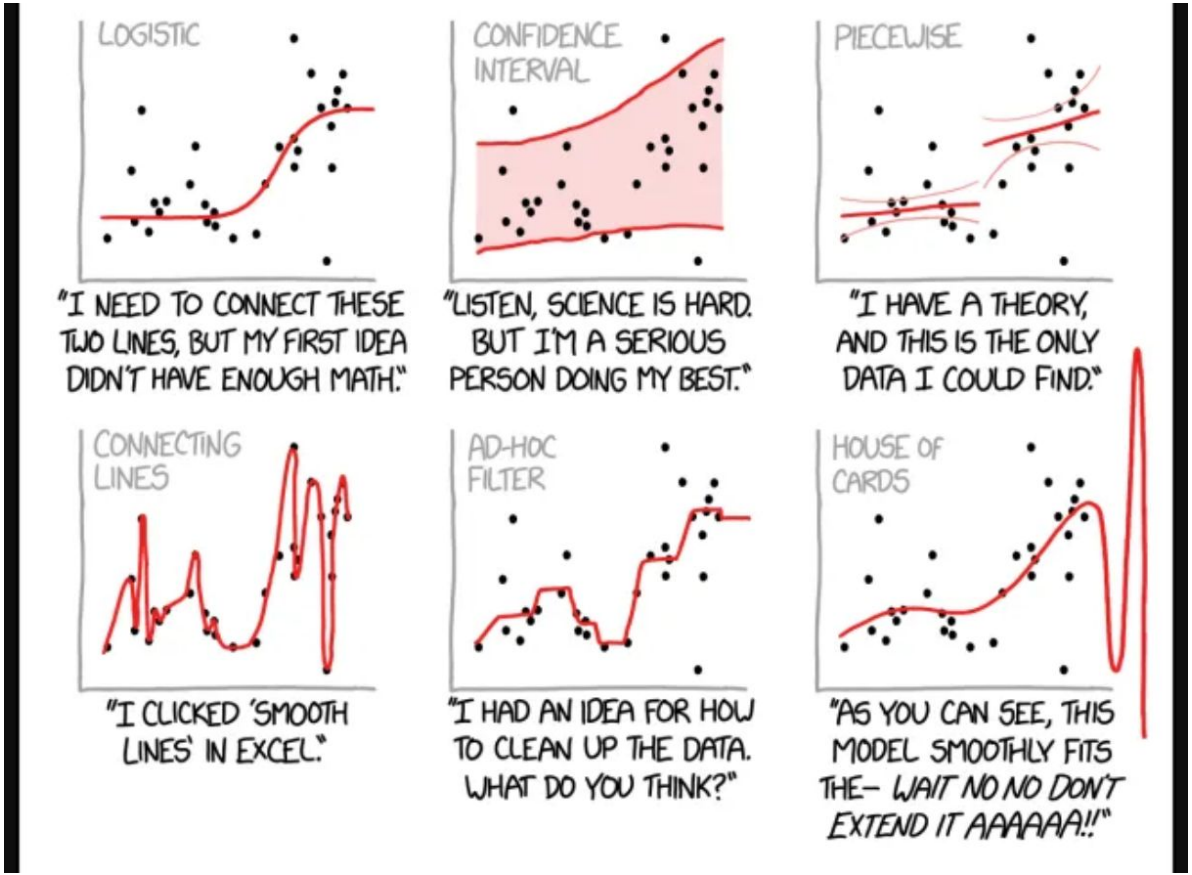
Cuadrados mínimos

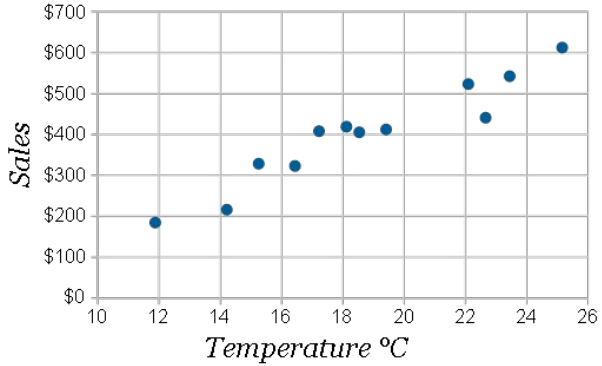
CURVE-FITTING METHODS
AND THE MESSAGES THEY SEND



Problema y Objetivo principal:

Buscamos encontrar modelos que ajusten y puedan predecir datos medibles del Universo.

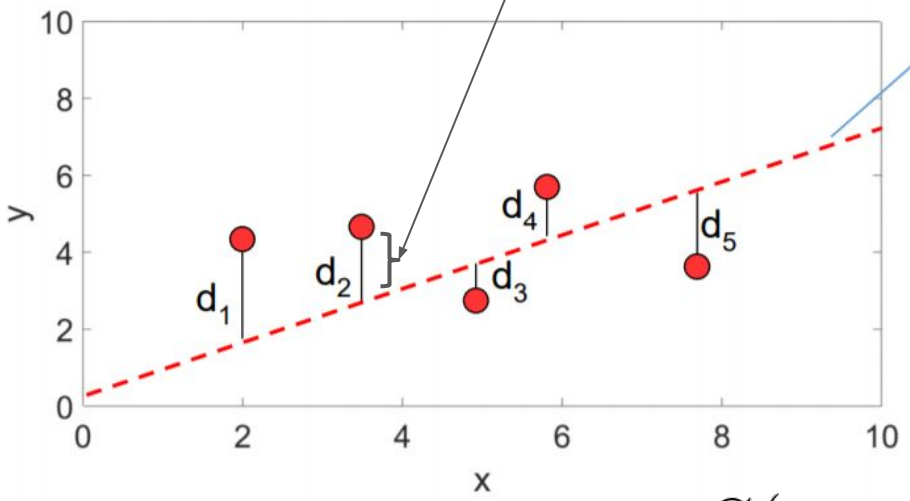
Partimos de lo fácil...



Tenemos una serie de mediciones (x_i,y_i), partimos de asumir una relación $y = f(x) = mx+q$ (modelo lineal)

Buscamos encontrar los parámetros m y q que minimicen la distancia entre datos y modelo

Residuos



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

¿Qué distancia?

$$d_1^2+d_2^2+d_3^2+d_4^2+d_5^2$$

(formalmente)

$$S(m,q)^2 = \sum [y_i - (mx_i + q)]^2$$

Buscamos minimizar la <u>suma de residuos cuadrada</u>

Cuenta formal

$$S(m,q) = \sum_{i=1}^{N} |(mx_i + q) - y_i|^2 = \sum_{i=1}^{N} y_i^2 + m^2 \sum_{i=1}^{N} x_i^2 + Nq^2 + 2mq \sum_{i=1}^{N} x_i - 2m \sum_{i=1}^{N} x_i y_i - 2q \sum_{i=1}^{N} y_i$$

$$\left.\begin{aligned} \frac{\partial S(m,q)}{\partial m} &= 0 \\ \frac{\partial S(m,q)}{\partial q} &= 0 \end{aligned}\right] \longrightarrow \left[\begin{aligned} 2m \sum_{i=1}^{N} x_i^2 + 2q \sum_{i=1}^{N} x_i - 2 \sum_{i=1}^{N} x_i y_i &= 0 \\ 2Nq + 2m \sum_{i=1}^{N} x_i - 2 \sum_{i=1}^{N} y_i &= 0 \end{aligned}\right] \longrightarrow$$

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

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

Encontramos la suma de cuadrados mínima

¿Y las incertezas?

- Solo consideramos las incertezas de y (asumimos son mayores*)
- Hagamos dos ajustes de cuadrados mínimos, uno considerando los errores y otro no.

Sin errores

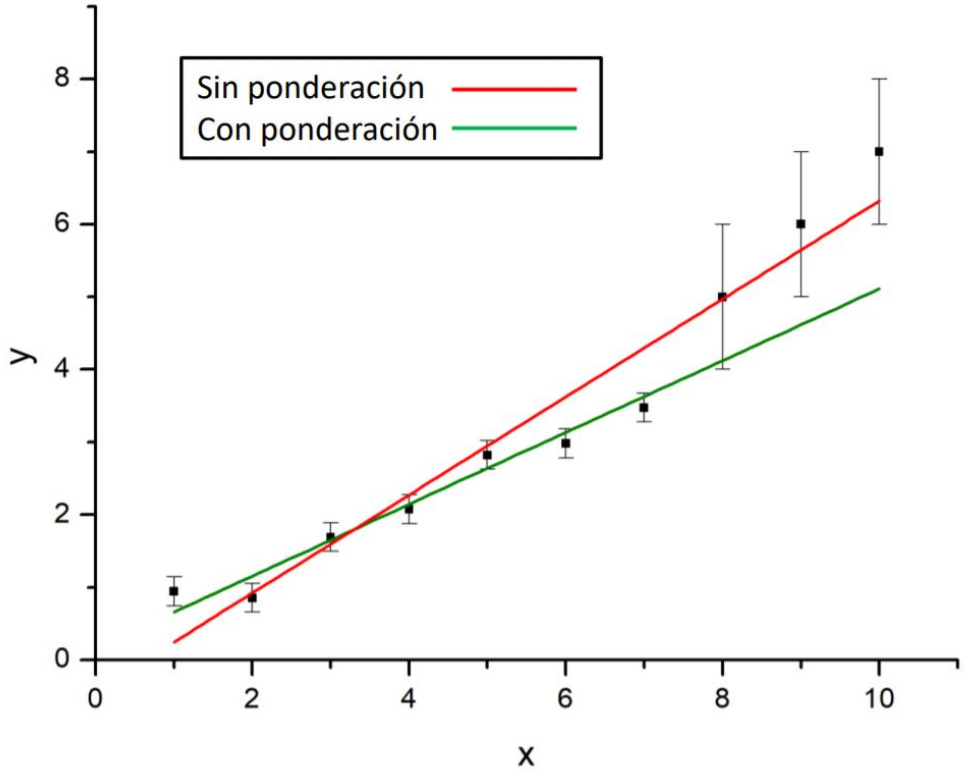
$$S^2 = \sum [y_i - (mx_i + q)]^2$$

$$\chi^2 = \sum [\frac{y_i - (mx_i + q)}{\sigma_i}]^2$$

Con errores

*Mayores implica: $m\Delta x >> \Delta y$

¿Y las incertezas?



Podemos ver claramente que el ajuste con ponderación (considerando errores de y) no dá tanta importancia a puntos con alto error.

$$S^2 = \sum [y_i - (mx_i + q)]^2$$

VS

$$\chi^2 = \sum [\frac{y_i - (mx_i + q)}{\sigma_i}]^2$$

¿Qué tipos de parámetros nos podemos encontrar?

$$\chi^2_\nu$$

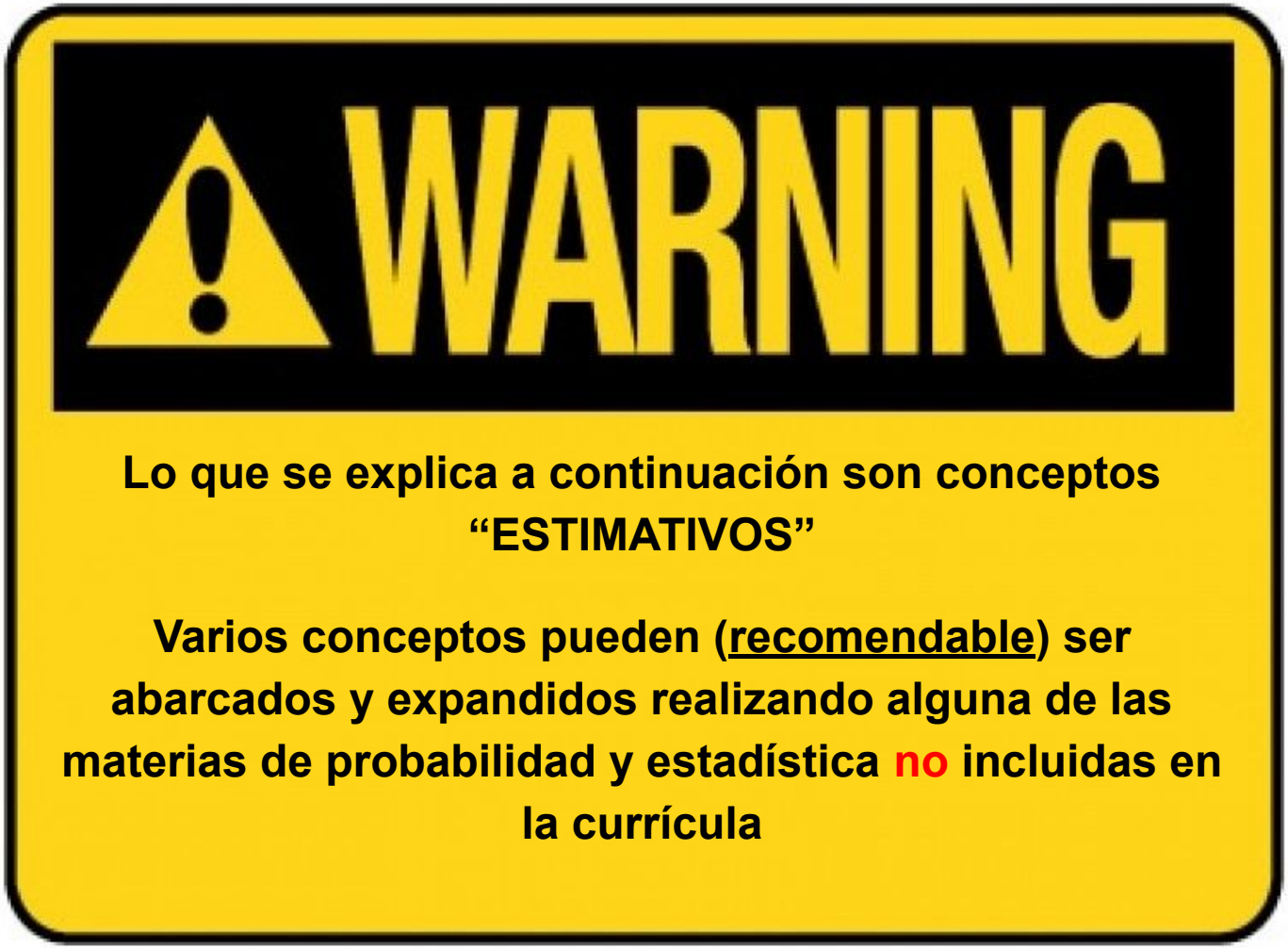
"Chi-cuadrado" (reducido)

Nos dá una noción de si nuestro **modelo** es "compatible" con los datos.
¿Quizás faltan parámetros?

$$R^2_{adj}$$

Coeficiente de determinación (ajustado)

Nos dá una noción de qué tan bueno fue el **ajuste** a los datos.


WARNING
Lo que se explica a continuación son conceptos "ESTIMATIVOS"
Varios conceptos pueden (recomendable) ser abarcados y expandidos realizando alguna de las materias de probabilidad y estadística no incluidas en la currícula

Coeficiente de determinación

Suma de cuadrados mínimos (residuos)

$$RSS = \sum (y_i - \widehat{y})^2$$

$$TSS = \sum (y_i - \overline{y})^2$$

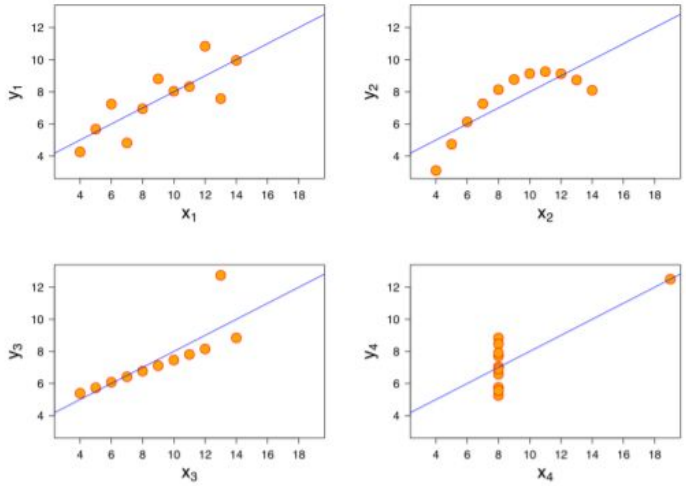
$$R^2 = 1 - \frac{RSS}{TSS}$$

Valor promedio $\overline{y}$

Valores ajustados $\widehat{y}$

R^2 Más cerca de 1 → “Mejor” ajuste.
¿Más cerca?

Ojo con el R^2 !!



Anscombe's quartet:

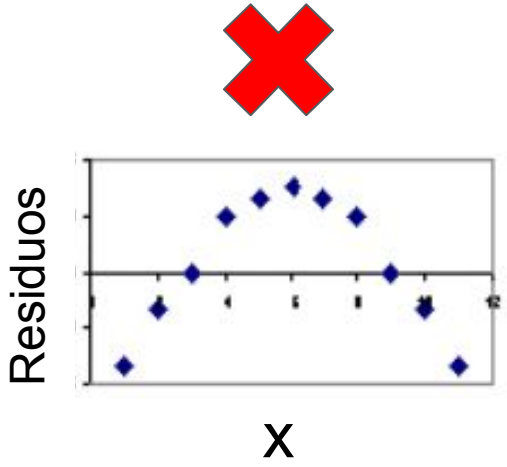
- Mismos valor medio y varianza de x.
- Casi mismos valores medios y varianzas de y.
- En los cuatro casos, mismo R^2

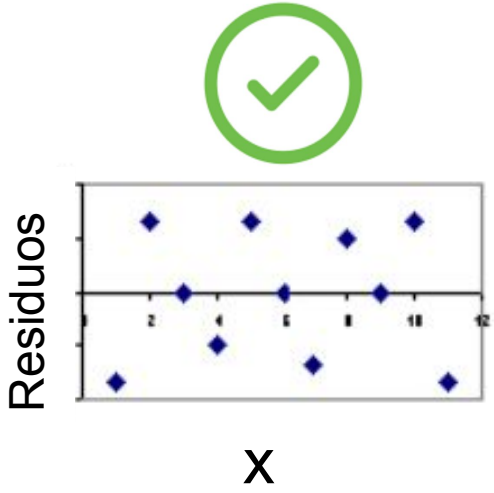
¿Qué está fallando?

Una cosa que está fallando es que la distribución de valores alrededor de la recta no es normal!

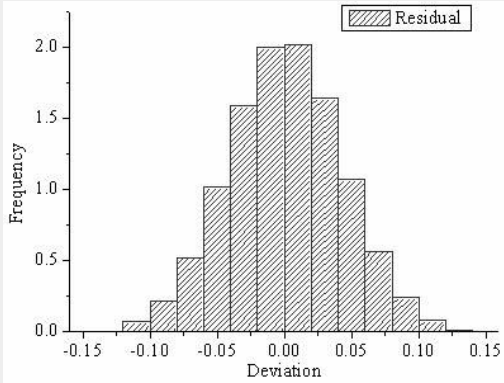
Podemos graficar los “residuos” (diferencia de cada punto con la recta ideal)

$$res_i = y_i - \hat{y}$$





Recordemos: Dist normal.



“Chi cuadrado”

$$\chi^2 = \sum \left[\frac{y_i - (mx_i + q)}{\sigma_i} \right]^2$$

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_k)^2}{(\sigma_i)^2}$$

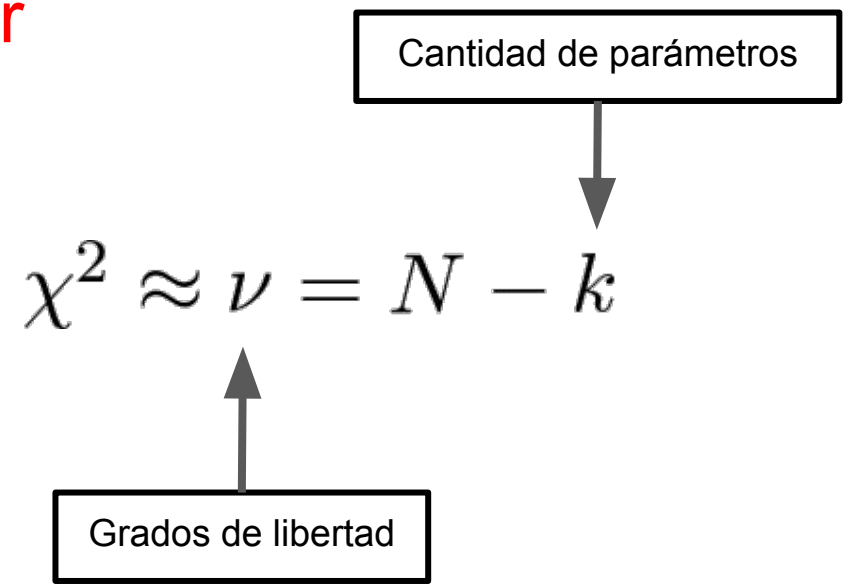
Oi: datos observados; Ek: datos ajustados

Tres casos para ajuste lineal:

N: número de datos

- $\chi^2 \simeq N - 2$: El modelo es presumiblemente compatible con el experimento.
- $\chi^2 \ll N - 2$: El modelo presumiblemente también es compatible con el experimento, pero las incertidumbres podrían estar sobreestimadas
- $\chi^2 \gg N - 2$: Probablemente, el modelo no ajuste a las mediciones

En gral (cualquier ajuste)



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

Chi cuadrado “reducido”:

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

- $\frac{\chi^2}{\nu} \sim 1$
- $\frac{\chi^2}{\nu} \ll 1$ 
- $\frac{\chi^2}{\nu} \gg 1$ 

Para resumir entonces (reglas **estimativas**)

Chi cuadrado **reducido** $\frac{\chi^2}{N-2}$

$$\chi^2_\nu$$

El modelo no es compatible

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

Incertezas sobreestimadas o datos ruidosos.

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

R cuadrado **ajustado**

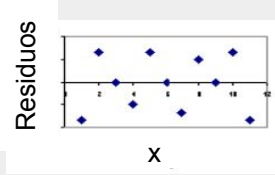
$$R^2_{adj}$$

Ajuste

$$R^2_{adj} \sim 1$$

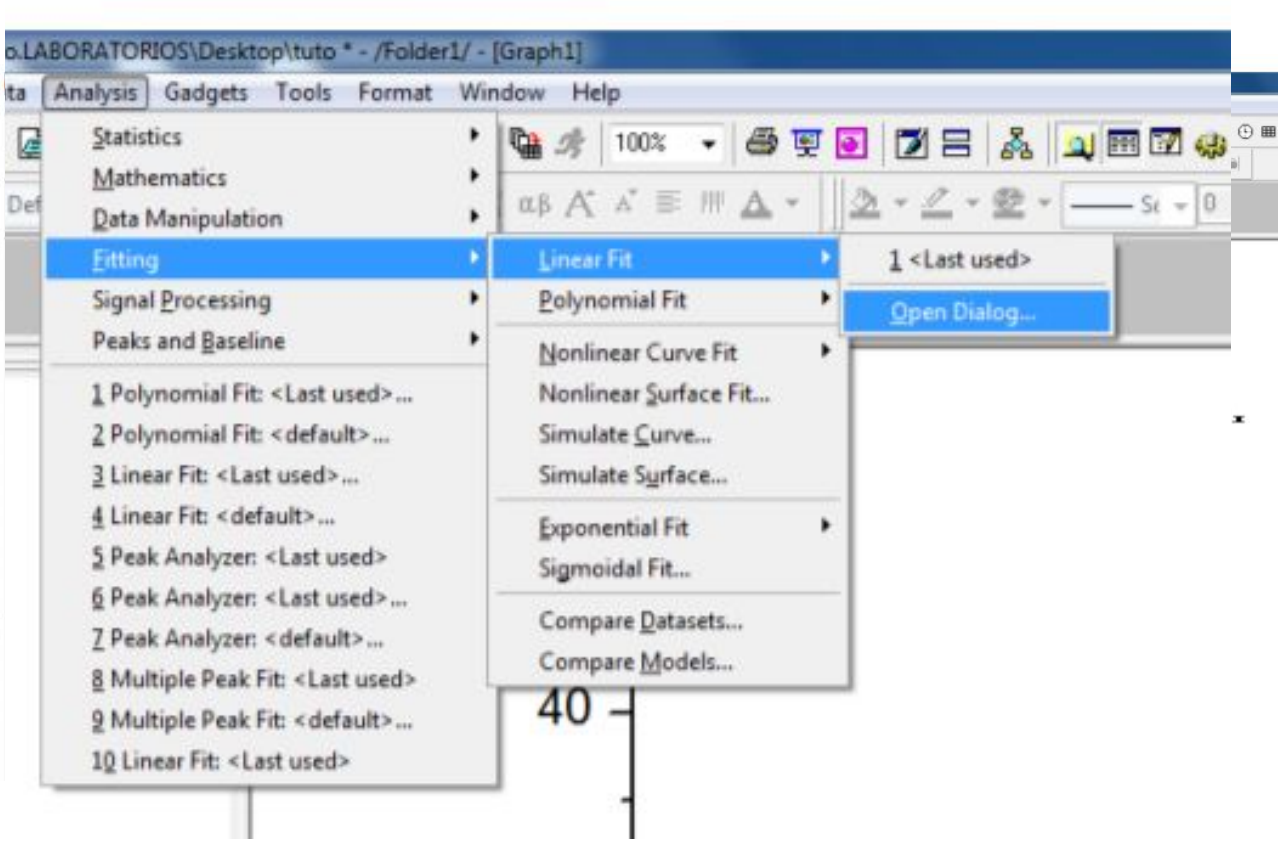
$$\chi^2_\nu \sim 1$$

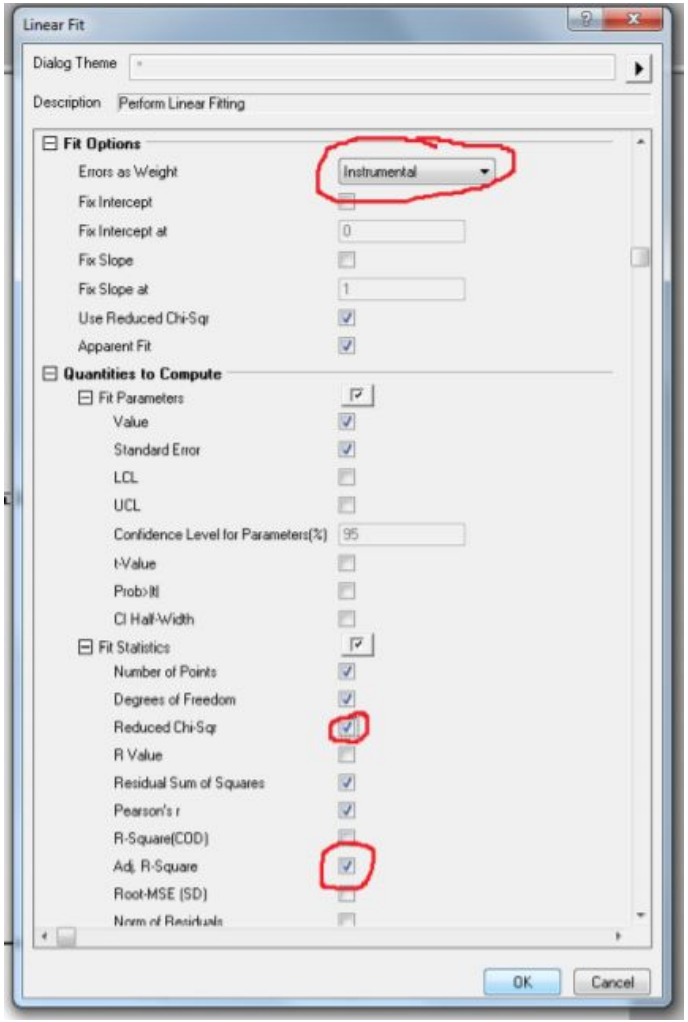
Análisis de residuos

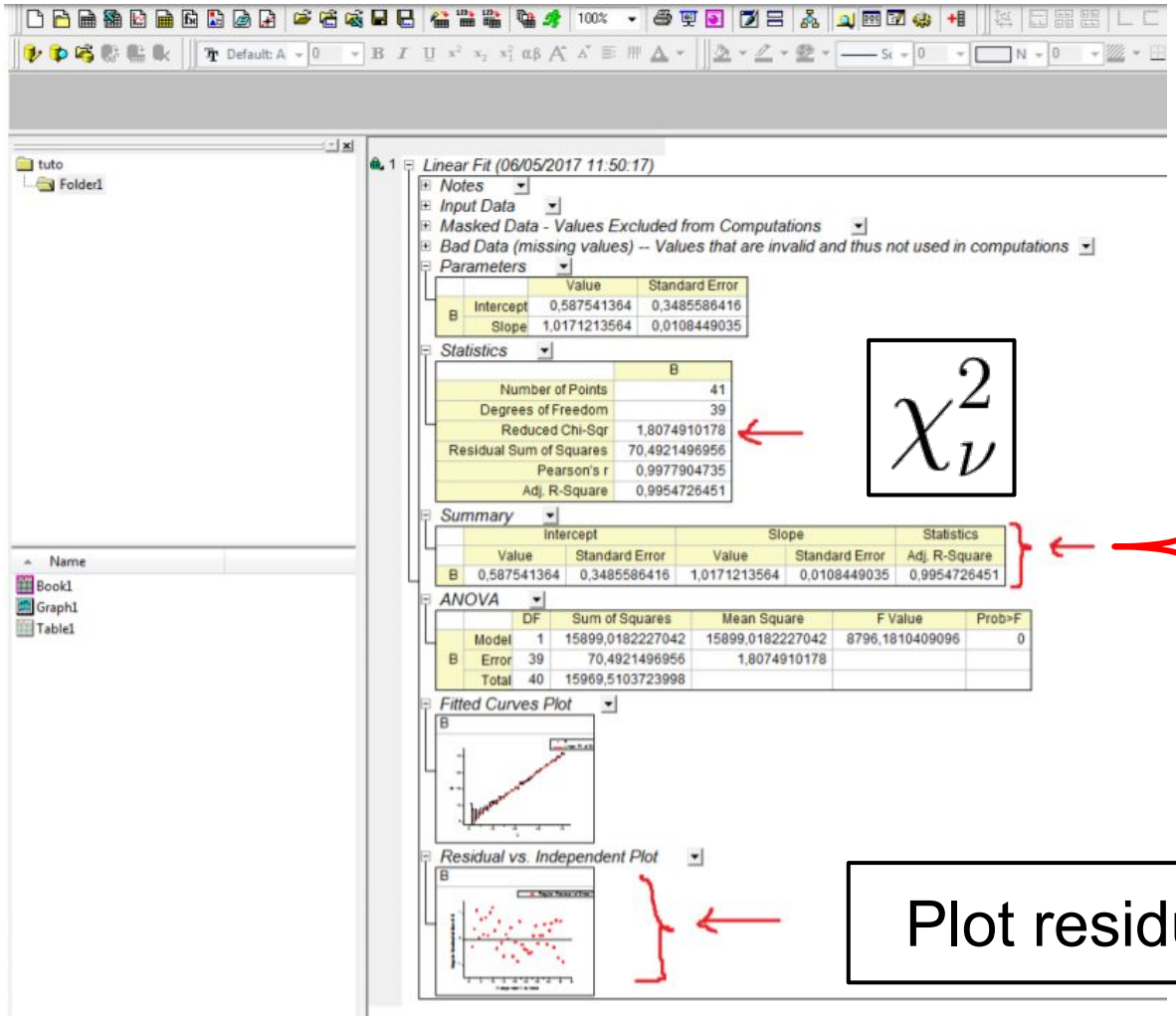


Finalmente, qué podemos usar en origin:

Analysis→
Fitting →
Linear Fit→
Open dialog...




Linear Fit
Dialog Theme
Description
Perform Linear Fitting
Fit Options
Errors as Weight
Instrumental
Fix Intercept
Fix Intercept at
0
Fix Slope
Fix Slope at
1
Use Reduced Chi-Sqr
Apparent Fit
Quantities to Compute
Fit Parameters
Value
Standard Error
LCL
UCL
Confidence Level for Parameters(%)
95
t-Value
Prob>|t|
CI Half-Width
Fit Statistics
Number of Points
Degrees of Freedom
Reduced Chi-Sqr
R Value
Residual Sum of Squares
Pearson's r
R-Square(COD)
Adj. R-Square
Root-MSE (SD)
Norm of Residuals
OK
Cancel



$$\chi^2_\nu$$

$$q \pm \Delta q$$
$$m \pm \Delta m$$
$$R^2_{adj}$$

Plot residual

Preguntas

Lectura adicional

https://www.originlab.com/doc/Origin-Help/Interpret-Regression-Result (para saber qué está haciendo origin atrás de cortinas)

https://www.originlab.com/doc/Origin-Help/Residual-Plot-Analysis (un poco de análisis de residuos)

https://cran.r-project.org/doc/contrib/Faraway-PRA.pdf (secciones 2.11 y 6.1)