Cuadrados Mínimos

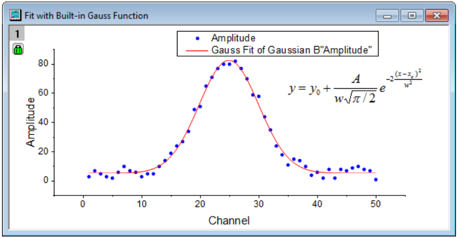
Diego Luna

May 8, 2017

Motivación

Dado un número de observaciones, puede requerirse el ajuste a un modelo

Ejemplo[1] :

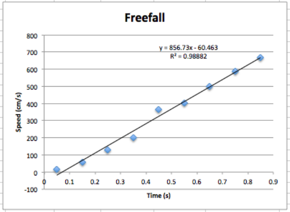


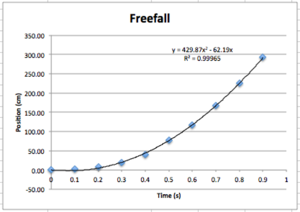
Parámetros: y_0, w, x_C, A
Curva roja: modelo
Puntos azules: mediciones

[1] http://www.originlab.com/index.aspx?go=Products/Origin/DataAnalysis/CurveFitting

Motivación

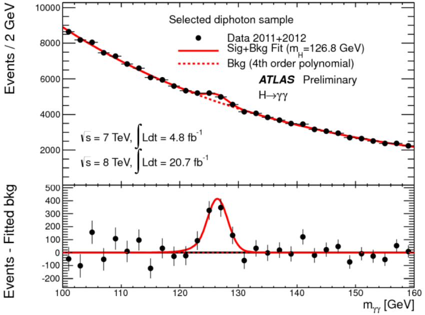
Ejemplo: Caida libre





Motivación

Ejemplo: Bosón de Higgs



Curva roja: modelo
Puntos negros: mediciones e incertidumbres

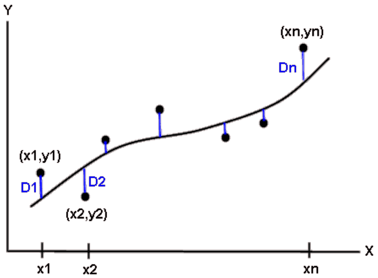
Usos

Se busca la función que mejor ajuste a los valores

Un buen proceso de ajuste debe incluir:

1. Parámetros de ajuste
2. Incertidumbres en los parámetros
3. Medida estadística de la bondad del ajuste

Cuadrados mínimos



Ejemplo: Tengo x_i, y_i, σ_i mediciones y un modelo $y(x, a_1, a_2, \ldots, a_M)$

- x_i, y_i : Mediciones
- σ_i : Incertidumbres de las mediciones y_i
- $y(x, a_1, a_2, \ldots, a_M)$: Modelo físico que quiero verificar
- $a_1, a_2, \ldots, a_M$: Parámetros que quiero determinar
- D_i : Distancia de la medición a la curva

Cuadrados mínimos

- La idea central en la técnica de cuadrados mínimos es minimizar la suma del cuadrado de las distancias a la curva.
- Pero ponderando según la incertidumbre de medición que tenga cada punto
- Vamos a estudiar el caso en que las incertidumbres en x son despreciables frente a las de y.

Ejemplo: Tengo x_i, y_i, σ_i mediciones y un modelo $y(x, a_1, a_2, \ldots, a_M)$

La mejor aproximación se obtiene de minimizar:

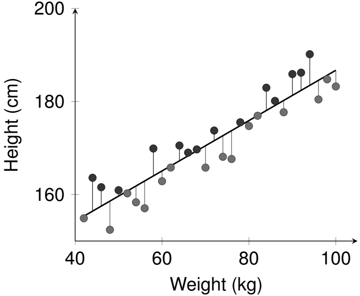
$$\chi^2 = \sum_{i=1}^{N} \left(\frac{D_i}{\sigma_i} \right)^2 = \sum_{i=1}^{N} \left(\frac{y_i - y(x_i, a_1, a_2, \ldots, a_M)}{\sigma_i} \right)^2 \tag{1}$$

Cuadrados mínimos: regresión lineal

Tengo N mediciones x_i, y_i, σ_i y un modelo $y = a + bx$

La mejor aproximación se obtiene de minimizar:

$$\chi^2(a,b) = \sum_{i=1}^{N} \left(\frac{y_i - a - bx_i}{\sigma_i} \right)^2 \tag{2}$$



Cuadrados mínimos: regresión lineal

Condición para minimización:

$$\frac{\partial \chi^2(a,b)}{\partial a} = 0 \qquad \frac{\partial \chi^2(a,b)}{\partial b} = 0$$

Definiciones:

- $S \equiv \sum \frac{1}{\sigma_i^2}$
- $S_x \equiv \sum \frac{x_i}{\sigma_i^2}$
- $S_y \equiv \sum \frac{y_i}{\sigma_i^2}$
- $S_{xx} \equiv \sum \frac{x_i^2}{\sigma_i^2}$
- $S_{xy} \equiv \sum \frac{x_i y_i}{\sigma_i^2}$
- $\Delta \equiv S S_{xx} - S_x^2$

Pendiente y ordenada

$$a = \frac{S_{xx} S_y - S_x S_{xy}}{\Delta} \tag{3}$$

$$b = \frac{S_{xy} S - S_x S_y}{\Delta} \tag{4}$$

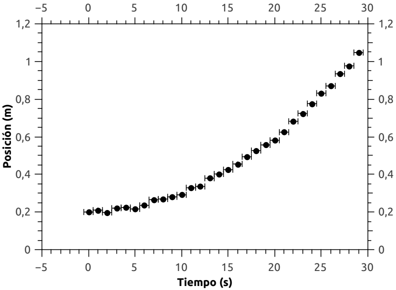
Cuadrados mínimos: regresión lineal

Incertidumbres en los parámetros

$$\sigma_a^2 = \frac{S_{xx}}{\Delta} \tag{5}$$

$$\sigma_b^2 = \frac{S}{\Delta} \tag{6}$$

Mediciones de la clase anterior



Tengo estas mediciones. A priori, puedo no conocer la dependencia:

- Modelo 1: $y(t) = Ae^{\alpha t}$ (exponencial)
- Modelo 2: $y(t) = At^{\alpha}$ (ley de potencia)

Mediciones de la clase anterior

Vamos a analizar el Modelo 2. Tareas:

1. Desplazar las mediciones para que partan del origen: $y_0 = 0$ y $x_0 = 0$
2. **Linealizar** los datos:
    - $y = \alpha t^{\beta} \Rightarrow ln(y) = ln(\alpha) + \beta ln(t)$
    - $ln(y) = ln(\alpha) + \beta ln(t)$
    - $ln(t) = \frac{1}{\beta} ln(y) - \frac{1}{\beta} ln(\alpha)$
3. Decidir qué valores van al eje vertical y cuáles al horizontal
4. Propagar incertidumbres. De Δy a $\Delta\left(ln(y)\right)$ ó de Δt a $\Delta\left(ln(t)\right)$
5. Calcular pendiente y ordenada al origen con sus incertidumbres: ¿Está dentro del valor esperado?
6. Agregar recta de ajuste en el gráfico

Cuadrados mínimos: regresión lineal

A calcular!

Cuadrados mínimos: regresión lineal

Coeficiente de correlación

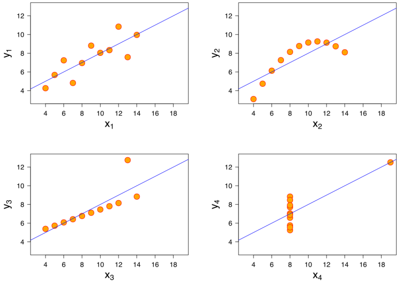
$$r \equiv \frac{\sigma_{xy}}{\sigma_x \sigma_y} \equiv \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2}\sqrt{\sum (y_i - \overline{y})^2}}$$

Características del estimador r:

- $-1 < r < 1$
- Suele usarse r^2
- Cuánto mas cerca de 1 esté r^2 , "mejor" (entre comillas)

Cuadrados mínimos: regresión lineal

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, $r = 0.816$

¿Qué está fallando?

Cuadrados mínimos: regresión lineal

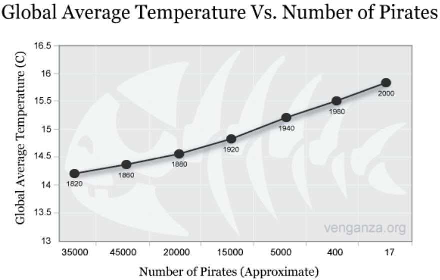
$\rightarrow$ La distribución alrededor de la recta no es normal.

Cuadrados mínimos: regresión lineal

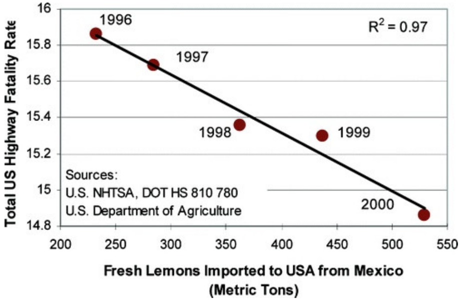
$\rightarrow$ La distribución alrededor de la recta no es normal.

Recordar:

- Siempre hay que graficar (y mirar) los valores
- **CORRELACIÓN NO IMPLICA CAUSALIDAD:**



Cuadrados mínimos: regresión lineal



Cuadrados mínimos: chi-test

La expresión de χ^2 puede escribirse como:

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_k)^2}{(\sigma_i)^2} \tag{7}$$

- O_i el valor observado (medición)
- E_i el valor esperado. Depende del modelo de medición que se tenga. Si es una recta, $E_i = a + bx_i$
- σ_i la incertidumbre en cada medición

Si uno hizo las cosas bien $\frac{(O_i - E_i)^2}{(\sigma_i)^2} \sim 1$
Para un ajuste a una recta, $\chi^2 \sim N - 2$ por los grados de libertad.

Cuadrados mínimos: chi-test

Tres casos para ajuste lineal:

- $\chi^2 \simeq N - 2$: 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

Cuadrados mínimos: chi-test

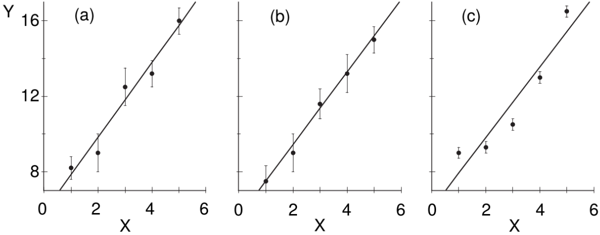
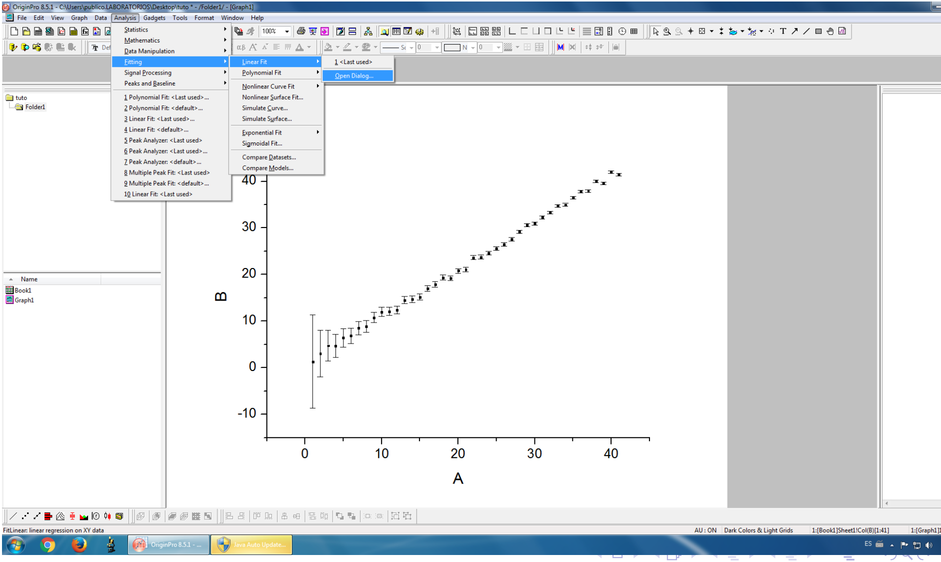


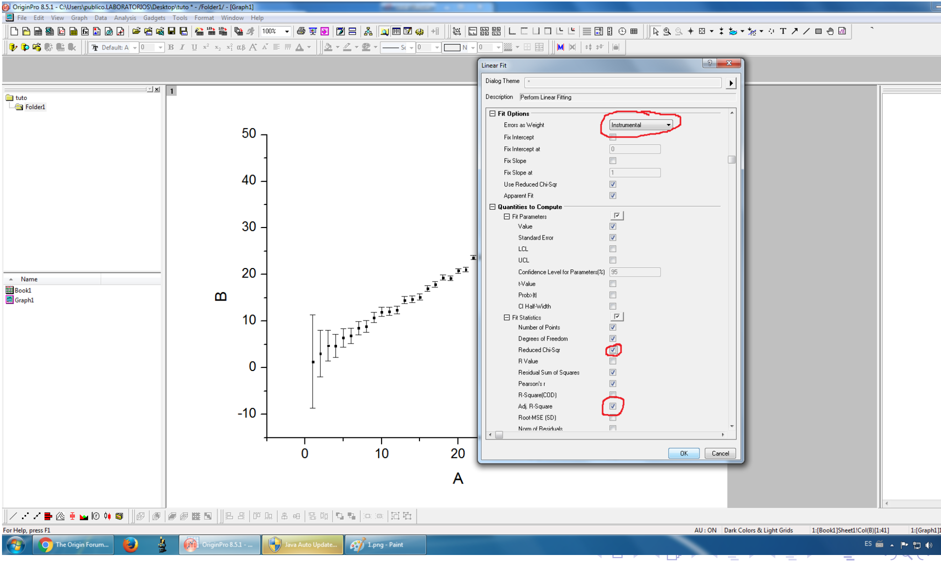
Fig. 11.1. The three graphs show three different situations relative to a linear fit to a set of data points: $\chi^2 \simeq \mathcal{N}$ (left), $\chi^2 \ll \mathcal{N}$ (center), $\chi^2 \gg \mathcal{N}$ (right). [2]

[2] Fornasini: The Uncertainty in Physical Measurements

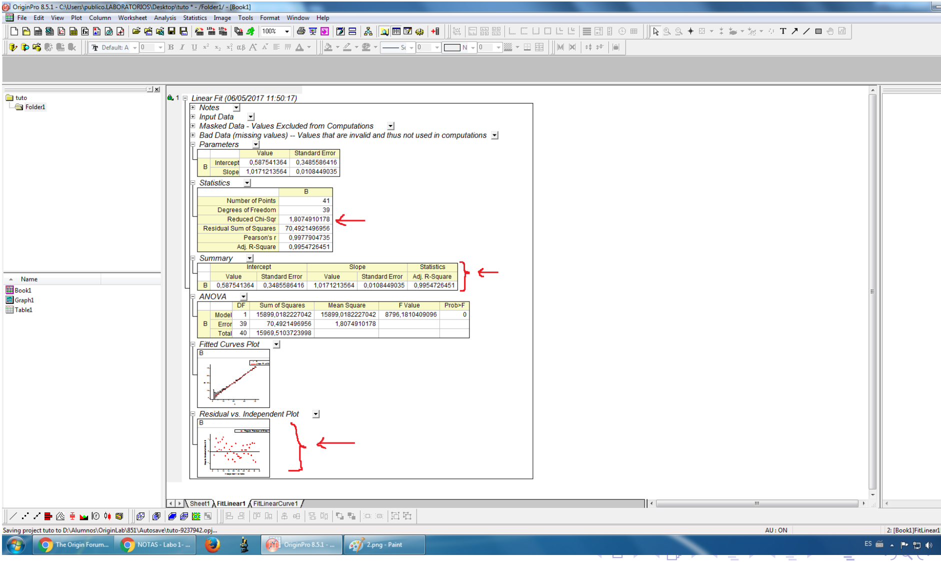
Cuadrados mínimos @Origin 8.5:



Cuadrados mínimos @Origin 8.5:



Cuadrados mínimos @Origin 8.5:



Referencias

Baird, D. C. (1962). Experimentation: an introduction to measurement theory and experiment design. Prentice Hall.

Fornasini, P. (2008). The uncertainty in physical measurements: an introduction to data analysis in the physics laboratory. Springer Science and Business Media.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1982). Numerical recipes in C (Vol. 2). Cambridge: Cambridge university press.