SOCIEDAD PERUANA DE FISICA SOPERFI FUNDADA EN 1975

Olimpiada Peruana de Física - 2012

Método de regresión lineal por mínimos cuadrados

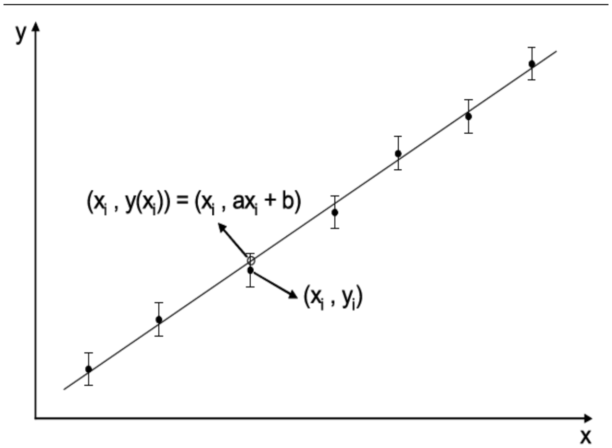
Vamos a suponer que tenemos um conjunto de N datos $(x_i,\ y_i)$, donde cada valor y_i tiene un error asociado que llamamos σ_i, o sea $(y_i \pm \sigma_i)$ (los σ_i no tienen que ser iguales). Vamos a suponer que los datos representan cierto fenómeno físico que cumple una ley descrita por uma función f.

Ajuste de una función lineal: Regreción lineal

Es el método de ajuste más adecuado para una distribución lineal. Sea un conjunto datos (x_i, y_i), las coordenadas del i-ésimo punto experimental, el subíndice i varía desde 1 (primer valor) hasta N (último valor).

$$f(x_i) = ax_i + b \tag{1}$$

Su representación gráfica sería



Si y_i es el valor medido, entonces,

$$ax_i + b \tag{2}$$

será el valor estimado, por lo que la desviación será

$$y_i - (ax_i + b). \tag{3}$$

Definiendo $w_i = 1/\sigma_i^2$, podemos escribir la desviación, tal que

$$s = \sum_{i=1}^{N} w_i \left[y_i - (ax_i + b)\right]^2 = \sum_{i=1}^{N} w_i \left[y_i - ax_i - b\right]^2 \tag{4}$$

sea mínima.
Aplicamos el método de mínimos cuadrados, para obtener a y b. Debe cumplirse las siguientes dos condiciones de mínimo de la función s.

$$\frac{\partial s}{\partial a} = 2 \sum_{i=1}^{B} w_i \left[y_i - ax_i - b\right](-x_i) = 0 \tag{5}$$

$$\frac{\partial s}{\partial b} = 2 \sum_{i=1}^{N} w_i \left[y_i - ax_i - b\right](-1) = 0 \tag{6}$$

Obtenemos, entonces, un sistema de dos ecuaciones con dos incógnitas (a y b)

$$\left(\sum_{i=1}^{N} w_i x_i^2\right) a + \left(\sum_{i=1}^{N} w_i x_i\right) b = \left(\sum_{i=1}^{N} w_i x_i y_i\right) \tag{7}$$

$$\left(\sum_{i=1}^{N} w_i x_i\right) a + \left(\sum_{i=1}^{N} w_i\right) b = \left(\sum_{i=1}^{N} w_i y_i\right) \tag{8}$$

Resolviendo el sistema de ecuaciones, los valores de a y b son:

$$a = \frac{\left(\sum_{i=1}^{N} w_i\right)\left(\sum_{i=1}^{N} w_i x_i y_i\right) - \left(\sum_{i=1}^{N} w_i y_i\right)\left(\sum_{i=1}^{N} w_i x_i\right)}{\Delta} \tag{9}$$

$$b = \frac{\left(\sum_{i=1}^{N} w_i y_i\right)\left(\sum_{i=1}^{N} w_i x_i^2\right) - \left(\sum_{i=1}^{N} w_i x_i y_i\right)\left(\sum_{i=1}^{N} w_i x_i\right)}{\Delta} \tag{10}$$

donde

$$\Delta = \left(\sum_{i=1}^{N} w_i\right)\left(\sum_{i=1}^{N} w_i x_i^2\right) - \left(\sum_{i=1}^{N} w_i x_i\right)^2 \tag{11}$$

Los errores asociados a a y b son:

$$\sigma_a^2 = \frac{\left(\sum_{i=1}^{N} w_i\right)}{\Delta} \qquad y \qquad \sigma_b^2 = \frac{\sum_{i=1}^{N} w_i x_i^2}{\Delta} \tag{12}$$

Estas ecuaciones son generales y valen para el caso en que los σ_i son diferentes.
En el caso en que los σ_i sean iguales ($= \sigma$, o sea, el mismo valor para todo i), las expresiones

de a, b, σ_a^2 y σ_b^2 se puden simplificar:

$$a = \frac{N\left(\sum_{i=1}^{N} x_i y_i\right) - \left(\sum_{i=1}^{N} y_i\right)\left(\sum_{i=1}^{N} x_i\right)}{\Delta} \tag{13}$$

$$b = \frac{\left(\sum_{i=1}^{N} y_i\right)\left(\sum_{i=1}^{N} x_i^2\right) - \left(\sum_{i=1}^{N} x_i y_i\right)\left(\sum_{i=1}^{N} x_i\right)}{\Delta} \tag{14}$$

donde

$$\Delta = N\left(\sum_{i=1}^{N} x_i^2\right) - \left(\sum_{i=1}^{N} x_i\right)^2 \tag{15}$$

Los errores asociados a a y b son:

$$\sigma_a^2 = \frac{\left(\sum_{i=1}^{N} w_i\right)}{\Delta} \qquad y \qquad \sigma_b^2 = \frac{\sum_{i=1}^{N} w_i x_i^2}{\Delta} \tag{16}$$

Una vez que se tienen los valores de a y b es necesario expresar de forma cuantitativa y cualitativa la calidad de ajuste. Por calidad se entiende intuitivamente lo que se separan los puntos experimentales y_i de la predicción de la recta $ax + b$. Cuantitativamente, la calidad de ajuste viene expresada numéricamente por el valor del coeficiente de correlación.
El **coeficiente de correlación** se define como

$$r = \frac{N\left(\sum_{i=1}^{N} x_i y_i\right) - \left(\sum_{i=1}^{N} y_i\right)\left(\sum_{i=1}^{N} x_i\right)}{\sqrt{N\left(\sum_{i=1}^{N} x_i^2\right) - \left(\sum_{i=1}^{N} x_i\right)^2}\sqrt{N\left(\sum_{i=1}^{N} y_i^2\right) - \left(\sum_{i=1}^{N} y_i\right)^2}} \tag{17}$$

Este coeficiente expresa la calidad del ajuste, en forma relativa es un número sin unidades. Puede demostrarse que cuanto más cercano a 1 sea el valor de r mejor será el ajuste y r nunca puede ser mayor que 1.