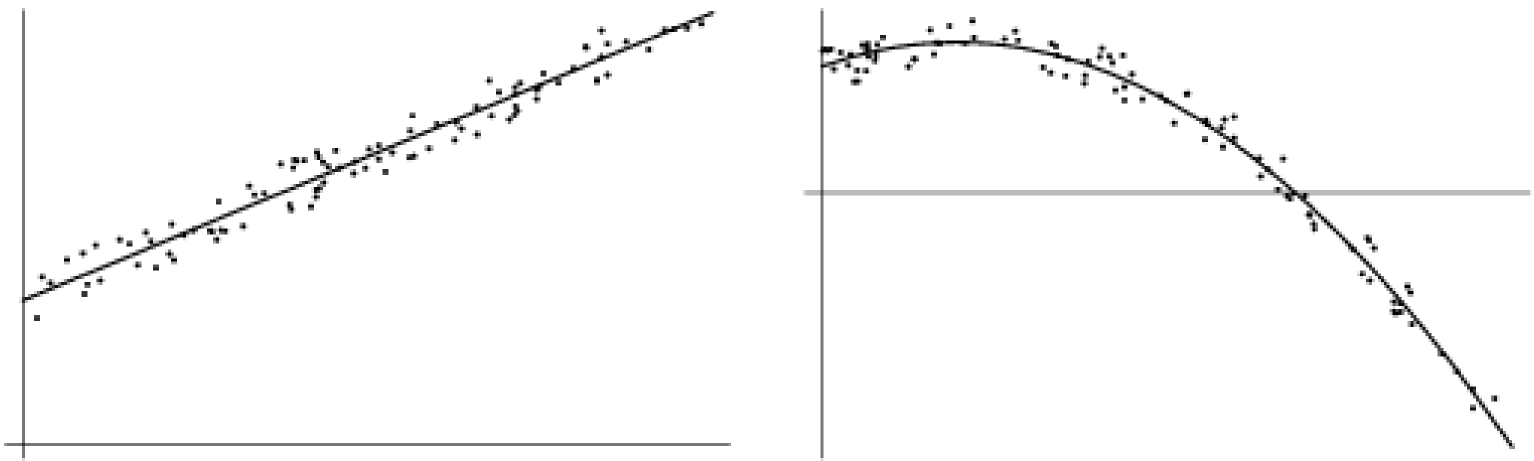
Cuadrados mínimos:

algoritmo (procedimiento) matemático que nos permite obtener la curva (función matemática, f) que “mejor ajuste” (modele, represente) a un conjunto dado de puntos.



¿Cómo se consigue?

Haciendo mínimo el apartamiento Di de los puntos a la curva

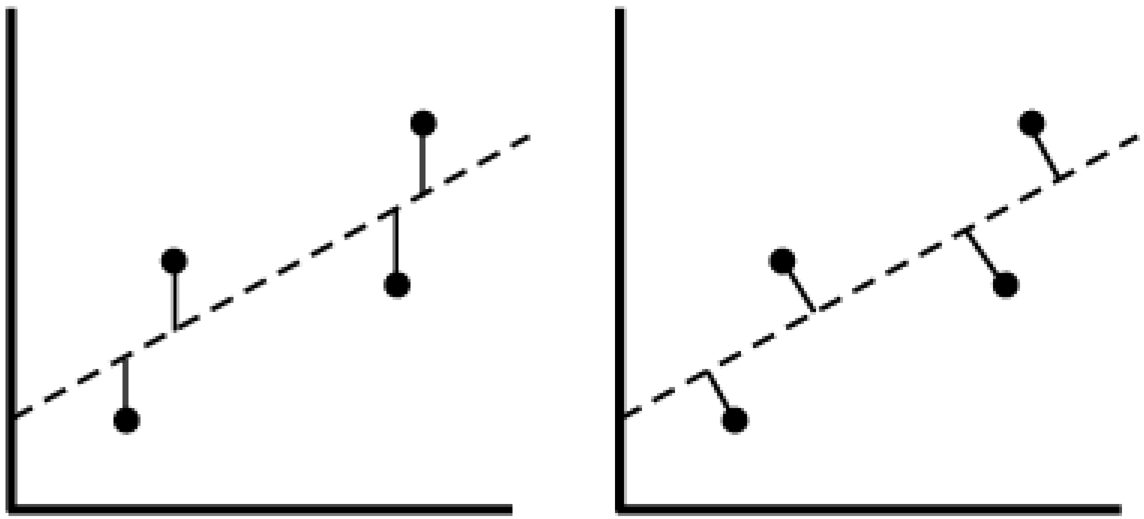
Supongamos una relación entre 2 variables. Para un punto de coordenadas x_i, y_i la función f tomará el valor $f(x_i)$ y

$$\boxed{D_i = y_i - f(x_i)}$$

Vamos a minimizar

$$\chi^2 = \sum D_i^2 = \sum_i \left[f(x_i) - y_i \right]^2$$

Porque es una función continua y diferenciable



Diferencia vertical

Diferencia perpendicular,

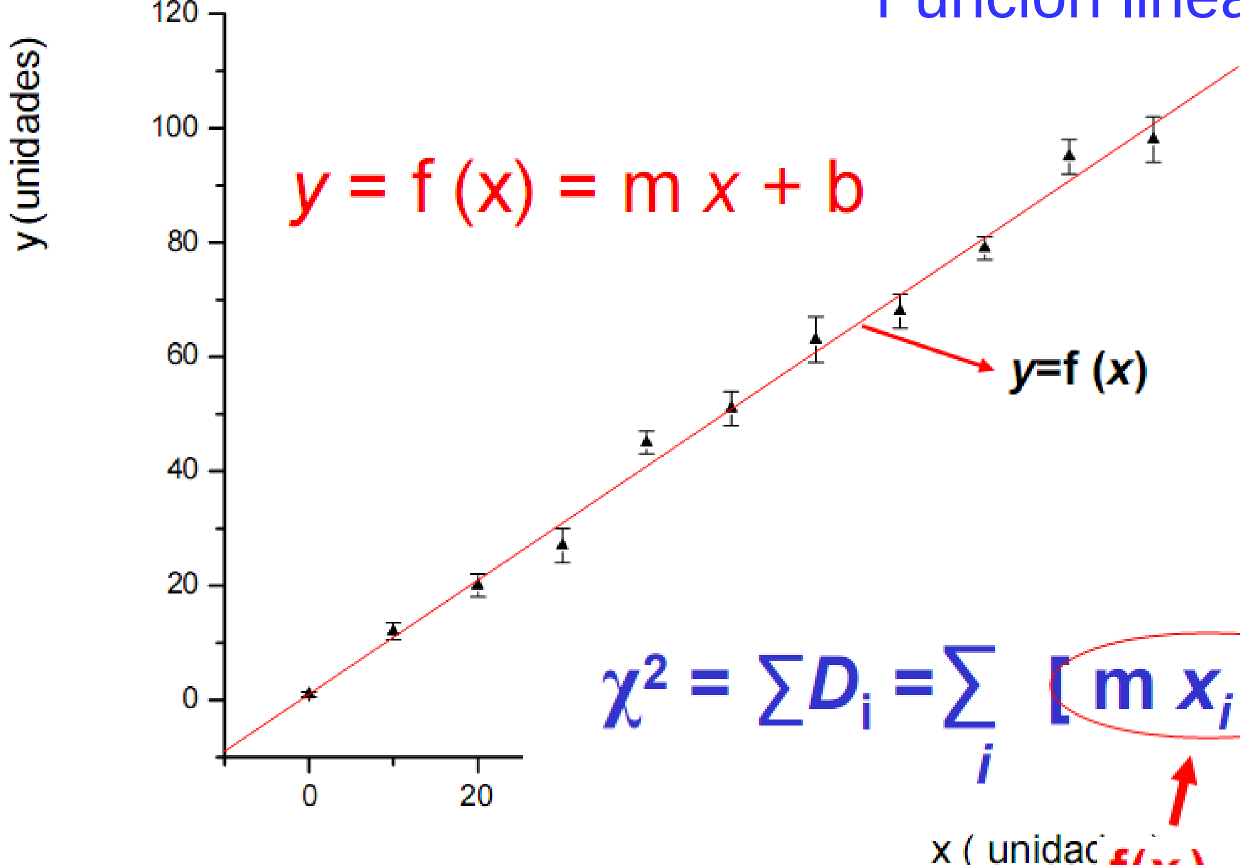
se trabaja con ésta porque las incertidumbres se pueden incorporar en forma simple

Para un conjunto de datos “ruidoso” la diferencia es pequeña

AJUSTE ANALÍTICO

Función lineal

$$y = f\,(x) = m\,x + b$$

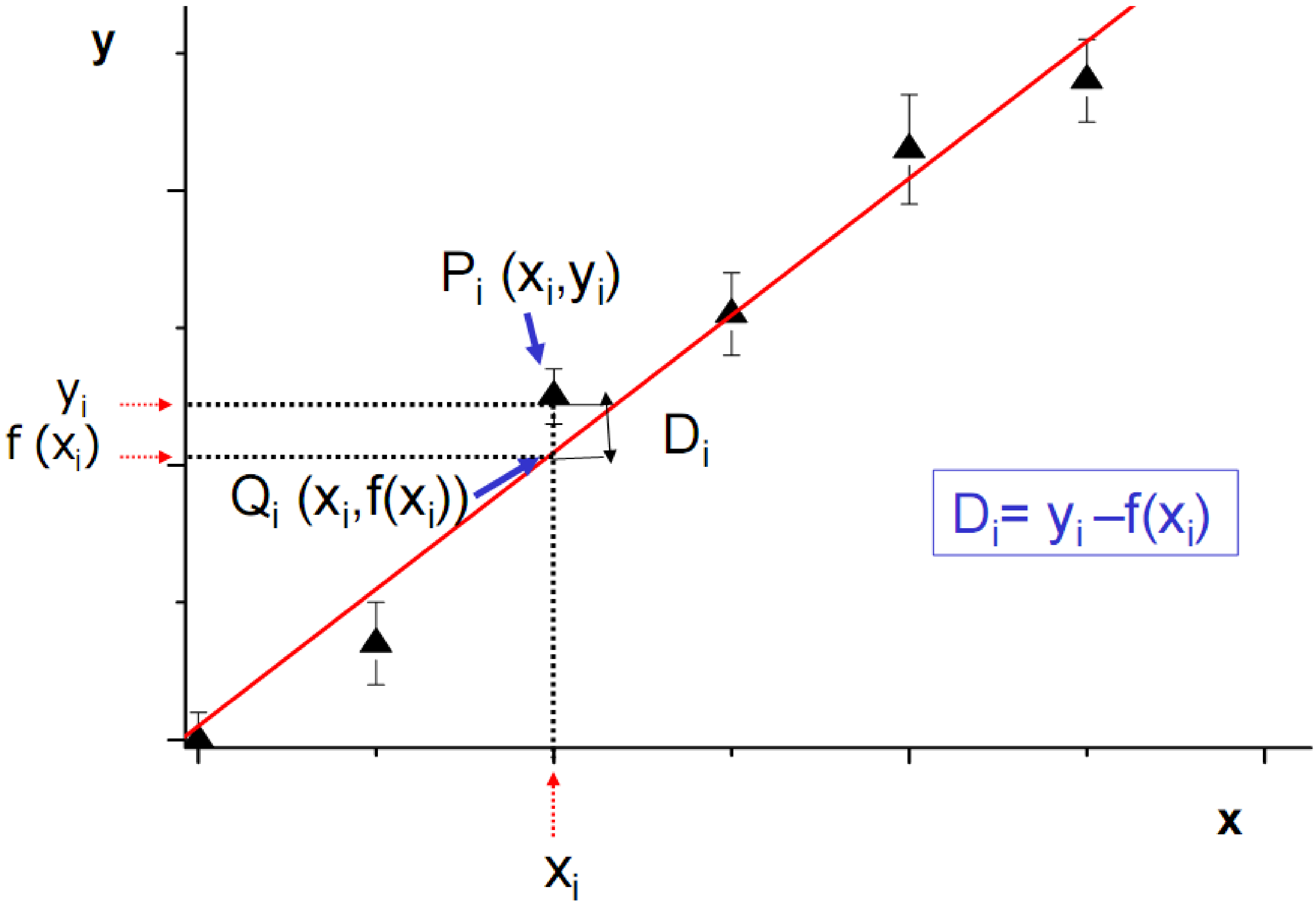


$$\chi^2 = \sum D_i = \sum_i \left[\, m\, x_i + b - y_i \right]^2$$

f(x_i)

Valor medido

Aumentando una zona del gráfico



¿Qué nos dice la Matemática sobre cómo encontrar el mínimo de una función?

La derivada primera de la función repecto de la variable tiene que ser cero y la segunda mayor que cero para el valor de la variable para el cual la función toma su mínimo valor

Es decir si tenemos una función f(x) que tiene un mínimo en A_o

$df(x)/dx|_{x=Ao} = 0$ y $d^2f(x)/dx^2|_{x=Ao} > 0$

Si se trata de un ajuste lineal para encontrar la mejor recta tenemos que encontrar los parámetros que definen la mejor recta o sea la pendiente y la ordenada al origen de la recta que minimice

$$\boxed{\chi^2 = \sum_i [y_i-(\mathbf{m}\, x_i + \mathbf{b})]^2}$$

Hay que encontrar el mínimo de una función de 2 variables

$$\boxed{\chi^2 = \sum_i [y_i - (m\, x_i + b)]^2}$$

$\partial\chi / \partial m = 0$ $\partial\chi / \partial b = 0$

Que nos llevan, si tenemos n puntos a :

$$m = [\,(\sum x_i)\,(\sum y_i) - n \sum x_i\, y_i\,] \,/\, [\,(\sum x_i)^2 - n \sum x_i^2\,]$$

$$b = [(\sum x_i)\,(\sum x_i\, y_i) - (\sum x_i^{\,2})(\sum y_i)] \,/\, [\,(\sum x_i)^2 - n \sum x_i^2\,]$$

Se pueden encontrar en forma exacta con una calculadora

¿Cuán bueno es el ajuste?

Necesitamos introducir otro conceptos

Covarianza

Correlación

covarianza

Da cuenta de cómo se relacionan la variación de dos variables por ej.

1. la capacidad pulmonar de un deportista con los días de entrenamiento
2. la capacidad pulmonar de un fumador con la cantidad de cigarrillo diarios

días de entrenamiento	Capacidad vital (cm^3)
0	4800
10	4840
20	4890
30 **<X> =30**	4930 **<y>=4982,86**
40	4980
50	5180
60	5260

N° de cigarrillos	Capacidad vital (cm^3)
0	4500
5	4200
10 **<x> = 10**	3300 **<y> = 3600**
15	3100
20	2900

En el caso 1 las dos variables covarian en la misma dirección,
en el 2 en direcciones opuestas.

covarianza y correlación nos vana permitir cuantificar
cómo covarian dos variables.

Si analizamos los datos de ambos conjuntos vemos que en el caso 1 cuando una de las variables toma un valor por encima del promedio la otra también y cuando la primera toma un valor por debajo del promedio la segunda también.

En el caso 2 cuando una de las variables toma un valor por encima del promedio la otra también toma un valor por debajo y viceversa.

Si promediamos el producto de las desviaciones de cada una podemos obtener información de cómo covarian ambas variables.

$$S_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)$$

Esta magnitud permanece invariante si a todo el conjunto de números se los suma por una número dado, pero no es invariante si se los multiplica por un número dado

Pasemos de $X_i\,, Y_i$ a $L_i\,, M_i$

$$L_i = aX_i + b,\; M_i = cY_i + d$$

$$\boxed{S_{LM}} = \frac{1}{N-1}\sum_{i=1}^{N} dl_i dm_i$$

$$= \frac{1}{N-1}\sum_{i=1}^{N} adx_i\, cdy_i = ac\frac{1}{N-1}\sum_{i=1}^{N} dx_i\, dy_i = \boxed{acS_{xy}}$$

covarianza

❖_Dado que las constantes multiplicativas aparecen en la

Covarianza un cambio de escala la modifica

❖ Es dificil de interpretar porque

trae incluída la información de la escala

Coeficiente de correlación

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

Medida estadística de cómo están relacionados los cambios de dos variables

$$-1 \leq r_{xy} \leq 1$$

Fórmulas para el cálculo

$$S_{xy} = \frac{1}{N - 1} \sum X_i Y_i - \overline{X}\,\overline{Y}$$

Covarianza (ver apéndice)

$$r_{xy} = \frac{N \sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{\left[N \sum X_i^2 - \left(\sum X_i \right)^2 \right] \left[N \sum Y_i^2 - \left(\sum Y_i \right)^2 \right]}}$$

Coeficiente de correlación

OJO correlación no quiere decir causa,

puede habar otros factores influyendo

Podría haber una tercer variable que modifique a las dos

que encontramos muy correlacionadas

Apéndice

$$
\begin{aligned}
s_{XY} &= \frac{1}{n}\sum_{i=1}^{n}\left(x_i y_i - \bar{x} y_i - x_i \bar{y} + \bar{x}\,\bar{y}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n}\bar{x} y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \bar{y} + \frac{1}{n}\sum_{i=1}^{n}\bar{x}\,\bar{y} \\
&= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\underbrace{\frac{1}{n}\sum_{i=1}^{n} y_i}_{\bar{y}} - \bar{y}\underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i}_{\bar{x}} + \frac{1}{n} n\bar{x}\,\bar{y} \\
&= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\,\bar{y} - \bar{x}\,\bar{y} + \bar{x}\,\bar{y} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\,\bar{y}.
\end{aligned}
$$