

Regresión lineal

Autor: Curso *Física por Ordenador*. Profesor Ángel Franco García, de la *Escuela Universitaria de Ingeniería Técnica Industrial de Eibar, España*. <http://www.sc.ehu.es/sbweb/fisica/default.htm>

En esta página, se describe el procedimiento de ajuste de los datos experimentales a una línea recta denominado regresión lineal, que se usa en el laboratorio en varias situaciones:

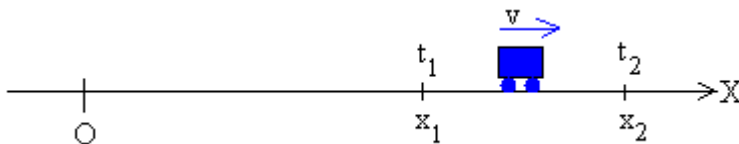
- Para calcular la velocidad en una experiencia de movimiento rectilíneo
- Para calcular la constante elástica de un muelle, colocando pesas en un platillo que cuelga de su extremo libre y midiendo la deformación del muelle
- etc.

El programa interactivo al final de esta página, está diseñado para que sea usado, en el Laboratorio de Física para cualquier experiencia que lo requiera. Nos proporciona los valores de:

- La pendiente a de la recta de regresión y el error cometido Δa
- La ordenada en el origen b
- El índice de correlación r . Este índice mide el grado de ajuste de los datos experimentales a la recta

Descripción

Supongamos que estamos midiendo la posición de un móvil en función del tiempo en un movimiento rectilíneo. Si el móvil está libre de fuerzas, esperamos que la relación entre la posición del móvil y el tiempo sea lineal $x=x_0+vt$. Donde x_0 es la posición del móvil en el instante $t=0$.



Si medimos las posiciones del móvil x_1 y x_2 en los instantes t_1 y t_2 , obtenemos un sistema de dos ecuaciones con dos incógnitas de las que podemos determinar las cantidades desconocidas x_0 y v . Ahora bien, esta afirmación solamente es cierta en un experimento ideal libre de errores.

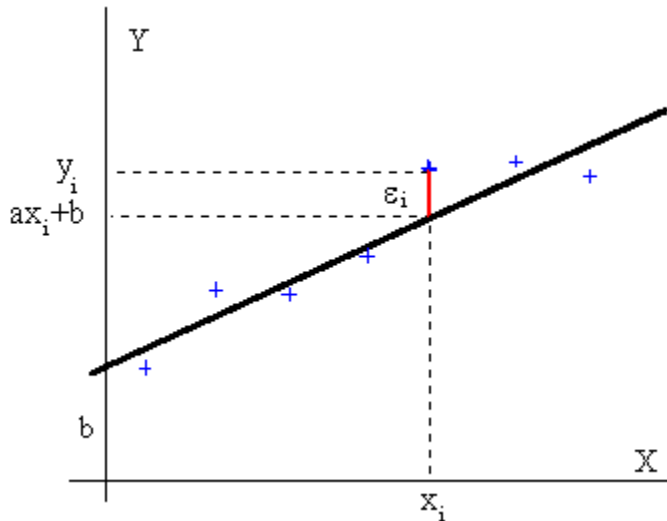
Si efectuamos n medidas de la posición del móvil, el aspecto de la representación gráfica de nuestras medidas puede ser parecido al de la figura más abajo, los puntos de color azul representan los datos experimentales. La relación entre las ordenadas y y las abscisas x de dichos puntos es solamente aproximada, debido a los errores de cada una de las medidas.

Si tomamos únicamente dos puntos para definir la recta el resultado tendría un importante error. Para una mejor estimación de la recta y por tanto, de las magnitudes buscadas, se deberán utilizar las n medidas tomadas.

Supongamos una magnitud física y , relacionada con otra x , mediante la función $y=ax+b$. Una recta de

pendiente a cuya ordenada en el origen es b . Las desviaciones ε de los valores de y , véase la figura, serán

- $\varepsilon_1 = y_1 - (ax_1 + b)$
- $\varepsilon_2 = y_2 - (ax_2 + b)$
-
- $\varepsilon_i = y_i - (ax_i + b)$
-
- $\varepsilon_n = y_n - (ax_n + b)$



Sea $E(a,b)$ la suma de los cuadrados de todas estas desviaciones

$$E(a,b) = (y_1 - ax_1 - b)^2 + (y_2 - ax_2 - b)^2 + \dots + (y_i - ax_i - b)^2 + \dots + (y_n - ax_n - b)^2$$

Los valores que minimizan a $E(a,b)$ son aquellos para los que

$$\frac{\partial E}{\partial a} = 0 \quad \text{y} \quad \frac{\partial E}{\partial b} = 0$$

Se obtiene así, un sistema de dos ecuaciones con dos incógnitas a y b cuya solución es

$$\boxed{\alpha = \frac{n \sum_1^n x_i y_i - \left(\sum_1^n x_i \right) \left(\sum_1^n y_i \right)}{n \sum_1^n x_i^2 - \left(\sum_1^n x_i \right)^2} \quad b = \frac{\sum_1^n y_i - \alpha \sum_1^n x_i}{n}}$$

Expresiones más elaboradas nos permiten determinar el error de a , Δa y el error de b , Δb

$$\Delta a = \frac{\sqrt{n} \cdot \sigma}{\sqrt{n \sum_1^n x_i^2 - \left(\sum_1^n x_i \right)^2}}$$

$$\sigma = \sqrt{\frac{\sum_1^n (y_i - ax_i - b)^2}{n - 2}}$$

$$\Delta b = \Delta a \cdot \sqrt{\frac{\sum_1^n x_i^2}{n}}$$

La pendiente de la recta se escribirá $a \pm \Delta a$, y la ordenada en el origen $b \pm \Delta b$. Véase las reglas para expresar una medida y su error de una magnitud.

El coeficiente de correlación es otro parámetro para el estudio de una distribución bidimensional, que nos indica el grado de dependencia entre las variables X e Y. El coeficiente de correlación r es un número que se obtiene mediante la fórmula.

$$r = \frac{\sum_1^n (x_i - \langle x \rangle)(y_i - \langle y \rangle)}{\sqrt{\sum_1^n (x_i - \langle x \rangle)^2} \sqrt{\sum_1^n (y_i - \langle y \rangle)^2}}$$

El coeficiente de correlación puede valer cualquier número comprendido entre -1 y +1.

- Cuando $r=1$, la correlación lineal es perfecta, directa.
- Cuando $r=-1$, la correlación lineal es perfecta, inversa
- Cuando $r=0$, no existe correlación alguna, independencia total de los valores X e Y