

El ajuste de datos por mínimos cuadrados

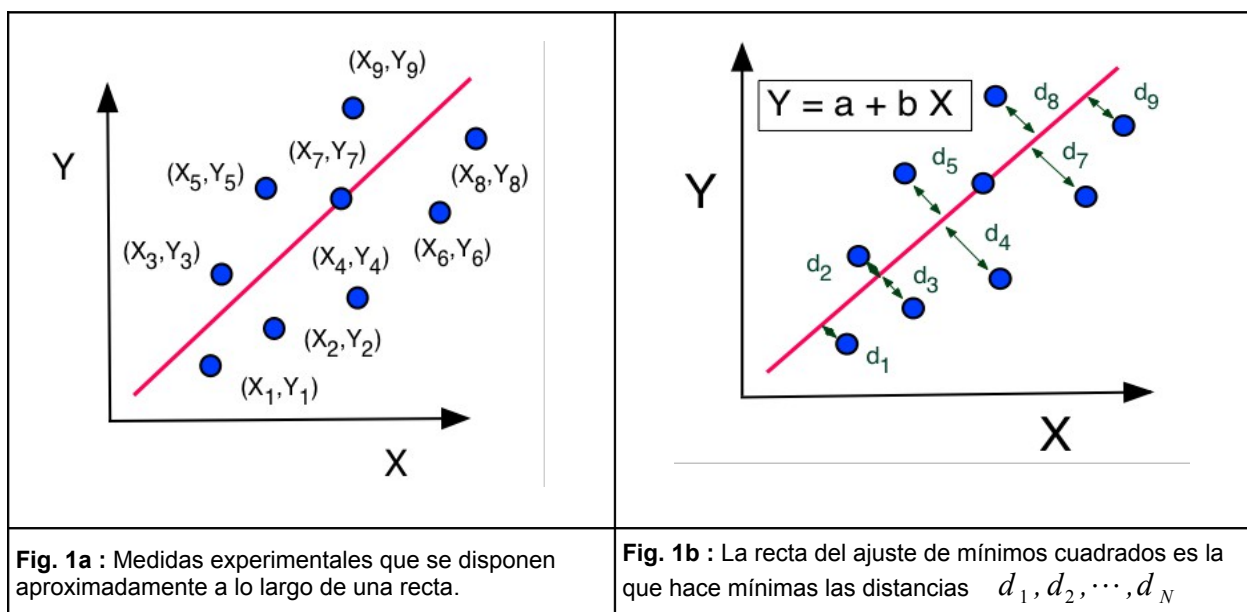
- **Finalidad** Introducir al estudiante mediante un ejemplo práctico al análisis de datos experimentales por el método de mínimos cuadrados.

En un laboratorio se comparan los resultados de los experimentos con fórmulas teóricas, es decir, comprobamos si las medidas concuerdan con los valores que predice la teoría. Como hemos visto, las mediciones siempre están afectadas por errores y para determinar con precisión el grado de concordancia es preciso efectuar cálculos estadísticos.

Esta tarea no siempre es fácil y en ciertos experimentos el análisis matemático de los datos puede ser muy complejo. Sin embargo, para los casos sencillos existen fórmulas de gran utilidad práctica, como es el *método de los mínimos cuadrados*. Aunque el caso más simple la función que ajustaremos será una recta, también se emplea para aproximar un conjunto de datos por algunas funciones más complejas. Por ejemplo, $Y = a e^{bX}$ o $Y = a X^b$ tomado logaritmos de los datos.

Aunque una teoría nos asegure que la relación entre dos magnitudes $Y = a + bX$ es lineal, en la práctica, la dispersión que causan los errores experimentales hace que no se encuentren *exactamente* dispuestos a lo largo de una recta⁽¹⁾ como muestra la figura 1a Siguiente. El método de mínimos cuadrados calcula a partir de los N pares de datos experimentales (X_i, Y_i) los valores de la pendiente b y del punto de corte con el eje a que mejor ajustan los datos a una recta.

La idea básica se muestra en la figura 1b, se entiende por *el mejor ajuste*, aquella recta que hace mínimas las distancias d_1, d_2, \dots, d_N de los puntos medidos a la recta.



No vamos a deducir aquí las fórmulas para la pendiente y el punto de corte que pueden encontrarse en cualquier texto de tratamiento de datos experimentales, sino que vamos a explicar cómo emplearlas mediante un ejemplo sencillo⁽²⁾. No son necesarios potentes ordenadores ni sofisticados método de cálculo, pues en el laboratorio de prácticas no se emplean más de 10-20 datos por práctica que pueden analizarse empleando una humilde calculadora. Para ilustrar el procedimiento, los gráficos y tablas del ejemplo de la presente nota se *han hecho a mano alzada deliberadamente* para mostrar la sencillez del procedimiento.

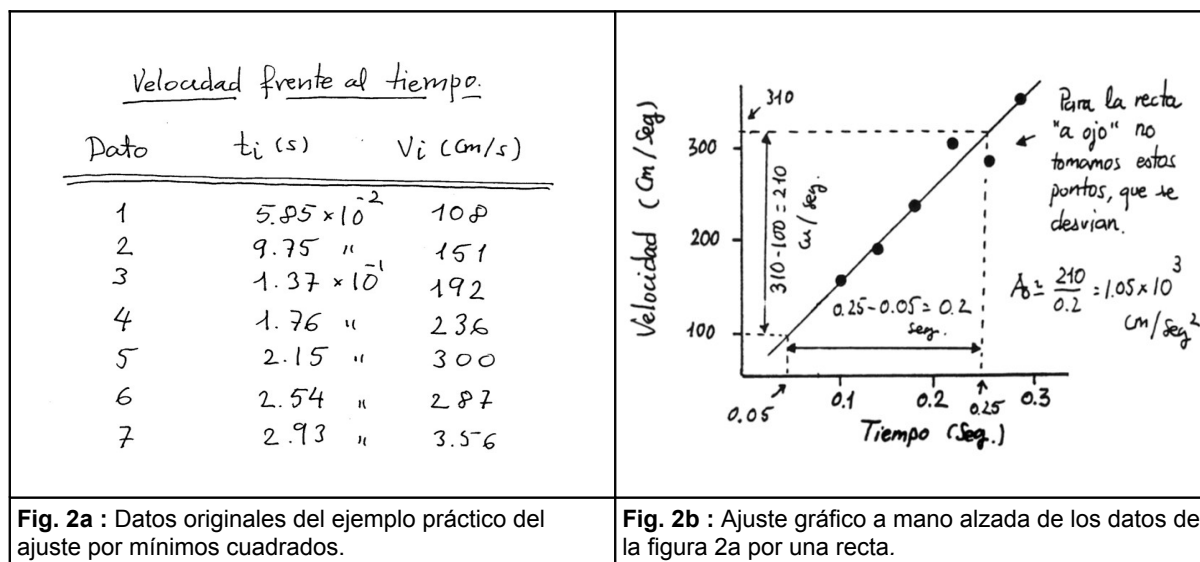
Las ecuaciones que hay que utilizar para ajustar mediante el método de mínimos cuadrados un conjunto de $i=1,2,\dots,N$ datos experimentales agrupados en parejas (X_i, Y_i) de medidas que deberían disponerse a lo largo de la recta $Y=a+b X$ son,

$$a = \frac{(\sum_{i=1}^N Y_i)(\sum_{i=1}^N X_i^2) - (\sum_{i=1}^N X_i)(\sum_{i=1}^N X_i Y_i)}{N(\sum_{i=1}^N X_i^2) - (\sum_{i=1}^N X_i)^2} \quad \text{y la pendiente} \quad b = \frac{N(\sum_{i=1}^N X_i Y_i) - (\sum_{i=1}^N X_i)(\sum_{i=1}^N Y_i)}{N(\sum_{i=1}^N X_i^2) - (\sum_{i=1}^N X_i)^2}$$

que aunque parecen complicadas pueden manejarse fácilmente si procedemos con orden.

Supongamos que hemos medido la velocidad de un cuerpo movido por una fuerza constante de modo que tendremos $V(t)=V_o+A t$ donde V_o es la velocidad inicial y $A=F_o/m$ su aceleración. Nuestras medidas, es decir, los datos de partida serán los siete pares de valores de figura 2a con los que queremos determinar V_o y A empleando las fórmulas anteriores.

Nuestras medidas originales serán las de la figura 2a siguiente. En este caso, X_i son los tiempos t_i y las velocidades $V_i(t_i)$ serán las coordenadas Y_i de la fórmula anterior.



En primer lugar es aconsejable dibujar los datos (por ejemplo, en un papel milimetrado) para comprobar si aproximadamente se disponen a lo largo de una recta como se muestra en la figura 2b puesto que si no fuese así, tendríamos que repetir las medidas.

Como puede observarse no es nuestro caso, pues si trazamos primero *una recta a ojo* con ayuda de una regla podemos ya tener un primer valor estimado de la pendiente. Su valor nos servirá para comprobar si hemos cometido un error en la manipulación de los datos.

La aceleración calculada con las fórmulas anteriores no puede ser muy diferente del valor aproximado de $A \approx 1,050 \times 10^3$ cm/s obtenido en la figura. Para emplear la fórmulas anteriores hay que calcular los siguientes sumatorios,

$$\sum_{i=1}^N X_i, \sum_{i=1}^N Y_i, \sum_{i=1}^N X_i^2 \text{ y } \sum_{i=1}^N X_i Y_i$$

Es fácil organizando los datos en columnas como muestra la figura 3. Las tres primeras son los datos departida anteriores y las dos nuevas los productos de cada par de datos que se suman y se ha indicado el resultado al final.

Cálculo de la aceleración por el método de mínimos cuadrados.

	X_i	Y_i	X_i^2	$X_i Y_i$
Dato num.	t_i (seg)	\bar{v}_i (cm/seg)	t_i^2	$\bar{v}_i \times t_i$
1	5.85×10^{-2}	108	3.42×10^{-3}	6.32
2	9.75 "	151	9.51	14.72
3	1.37×10^{-1}	192	1.88×10^{-2}	26.30
4	1.76 "	236	3.10 "	41.54
5	2.15 "	300	4.62 "	64.50
6	2.54 "	287	6.45 "	72.90
7	2.93 "	356	8.59 "	104.30
Sumas $N=7$	1.23	1.63×10^3	0.259	3.31×10^2

Fig. 3: Los datos ordenados en columnas para efectuar el cálculo del ajuste por mínimos cuadrados.

Con estos datos tenemos finalmente,

$$a = \frac{7 \times (3.31 \times 10^2) - (1.23) \times (1.63 \times 10^3)}{7 \times 0.259 - (1.23)^2} = 1.07 \times 10^3$$

que es un valor muy aproximado al obtenido mediante el análisis gráfico y dejamos que el lector calcule el valor del punto de corte con el eje.

Este procedimiento puede extenderse a un conjunto tan grande de datos como sea necesario. Aunque hemos mostrado como hacer los cálculos *a mano*, el trabajo se simplifica mucho empleando una hoja de cálculo. Existen programas de representación y análisis de datos experimentales que facilitan esta tarea y que disponen de otros algoritmos de aproximación mucho más complejos y sofisticados que el método expuesto.

Un último comentario se refiere al error cometido en cálculo del valor de la pendiente y el punto de corte del ajuste por mínimos cuadrados empleando las fórmulas anteriores. Como se ha hecho énfasis a lo largo del curso, *toda magnitud tiene un error absoluto y relativo en consecuencia a y b también*. Sin embargo, calcular dicho error es más complicado que *a* y *b*, por lo en este curso eximimos a los alumnos de evaluarlo. Sólo efectuaremos comparaciones de

la pendiente y del punto de corte de los ajustes cuando exista un valor teórico conocido, simplemente para comprobar cuánto se desvía del mismo.

Podemos hacernos una idea de lo complicado que es si observamos que los puntos de la Fig. 1a deben tener cada uno de ellos un error (no dibujado en la figura por claridad). Evidentemente, el peso estadístico de cada uno de ellos ha de tener en cuenta su error, que no ha de ser necesariamente igual para todos. En consecuencia, las distancias de la Fig. 1b no tienen el mismo peso dificultando considerablemente el cálculo.

Notas:

- (1) No hemos dibujado en las figuras las barras de error de los datos en este caso por simplicidad.
- (2) Muchos de las hojas de cálculo comerciales (*Excel*, *Numbers*, ...etc) y programas de análisis de datos (*Origin*, *ProFit*, *SciDAVis*, ...etc) incorporan la posibilidad de aproximar por mínimos cuadrados un conjunto de datos. Estos programas ponen a disposición de sus usuarios muchos más métodos de aproximación, algunos muy sofisticados. Por nuestra parte, no existe inconveniente si el alumno desea emplearlos siempre que comprenda lo que está haciendo y no se limite a obtener un número. La experiencia demuestra que una manipulación incorrecta de estos sofisticados códigos hace que los alumnos cometan errores que dan lugar a resultados disparatados.