

### 1.1. Regresión lineal

En la sesión práctica asociada a esta guía se estudiarán las resonancias de modos de flexión en una barra de Duraluminio mediante la técnica de impulsión. Usando una estimación del Módulo de Young se determinarán las frecuencias de algunos modos de flexión. Realizando una regresión lineal en escala log-log de estas frecuencias versus el número  $n$  de cada modo se obtendrá una medición precisa del Módulo de Young del material utilizado.

Es necesario revisar entonces la metodología de regresión lineal de un conjunto de datos. En la figura 1 se presenta una serie de medidas (puntos rojos) de una cierta cantidad  $y$  en función de otra cantidad  $x$ . Supondremos que se disponen de  $N$  datos, de modo que los resultados a analizar serán denotados como  $(x_i, y_i)$  con  $i = 1 \dots N$ . En la misma figura se muestra además una línea continua que pretende describir los resultados con el modelo lineal  $y = a \cdot x + b$ . La pregunta fundamental entonces que nos conierne es cómo determinar los mejores valores  $a$  y  $b$  de nuestro modelo que describan las medidas obtenidas.

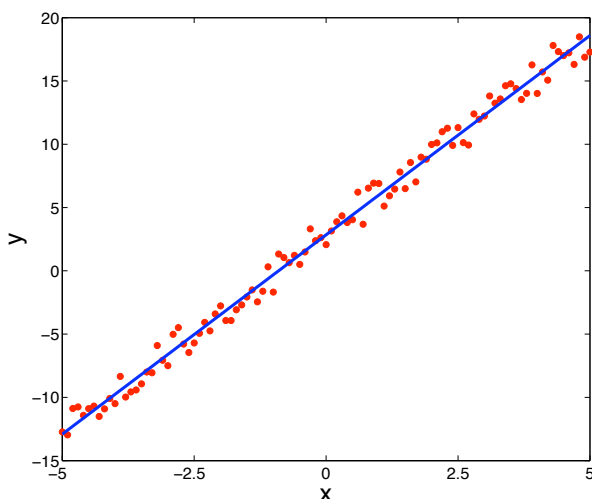


Figura 1: Ejemplo de conjunto de datos a los cuales se desea ajustar un modelo lineal  $y = a \cdot x + b$ . El método de regresión lineal permite encontrar los mejores valores  $a$  y  $b$  que se ajustan a los datos experimentales como también entregar una medida de lo bueno que es el ajuste mediante el parámetro  $R^2$ . En este caso los parámetros ajustados dan  $a = 3,15$  y  $b = 2,84$ , con  $R^2 = 0,992$ .

Para ello se define  $y_i^m = a \cdot x_i + b$  como el valor de la cantidad  $y$  dado por nuestro modelo evaluado en la cantidad  $x_i$ . Esto corresponde a los puntos sobre la curva continua de la figura 1 evaluados en los puntos  $x_i$ . Luego, en general, se cumple  $|y_i^m - y_i| \neq 0$  para todo  $i = 1 \dots N$ .

La búsqueda de los mejores valores de  $a$  y  $b$  se traduce entonces en minimizar estas cantidades  $|y_i^m - y_i|$ . Para ello se define

$$\chi^2 = \sum_{i=1}^N (y_i^m - y_i)^2, \quad (1)$$

la cual se busca minimizar. En el caso del modelo lineal se tiene

$$\chi^2 = \sum_{i=1}^N (a \cdot x_i + b - y_i)^2. \quad (2)$$

Para minimizar esta cantidad podemos comenzar por dar una interpretación geométrica. Para un conjunto de datos  $(x_i, y_i)$  se puede considerar  $\chi^2$  como una función de  $a$  y  $b$ , lo que puede ser considerado como una superficie: en el plano  $(a, b)$  la posición de la superficie corresponde al valor de  $\chi^2$ . Un ejemplo de esta interpretación se presenta en la figura 2a. Luego, el minimizar  $\chi^2$  se traduce en buscar los valles de esta superficie. La figura 2b muestra la coordenada de este mínimo (intersección de dos líneas punteadas).

Matemáticamente se obtiene el par de valores  $a$  y  $b$  que minimizan  $\chi^2$  imponiendo las condiciones

$$\frac{\partial \chi^2}{\partial a} = \frac{\partial \chi^2}{\partial b} = 0. \quad (3)$$

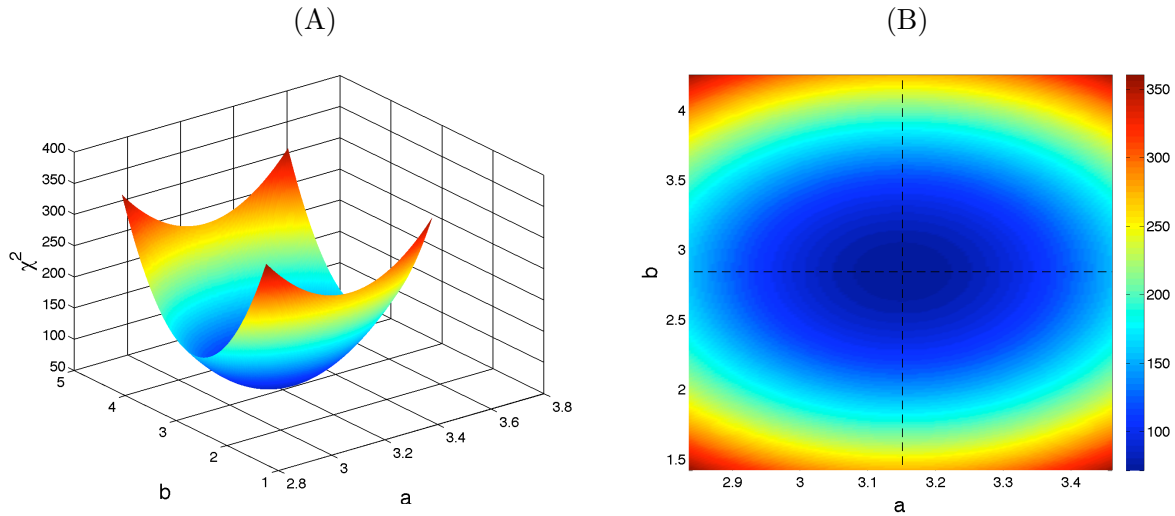


Figura 2: Superficie  $\chi^2$  versus  $a$  y  $b$  usando los datos  $(x_i, y_i)$  de la figura 1. El mínimo se obtiene para  $a = 3,15$  y  $b = 2,84$ , con  $R^2 = 0,992$ . La figura (A) muestra la superficie en una representación 3D y la figura (B) muestra la misma pero en una representación de colores, siendo azul oscuro el mínimo. La intersección de las líneas punteadas muestra el mínimo de  $\chi^2$  en  $a = 3,15$  y  $b = 2,84$ .

Luego, usando la definición (2), se obtiene el par de ecuaciones

$$a \sum_{i=1}^N x_i^2 + b \sum_{i=1}^N x_i - \sum_{i=1}^N x_i \cdot y_i = 0, \quad (4)$$

$$a \sum_{i=1}^N x_i + b \sum_{i=1}^N 1 - \sum_{i=1}^N y_i = 0. \quad (5)$$

Finalmente se puede despejar  $a$  y  $b$ :

$$a = \frac{\left( \frac{\sum_{i=1}^N x_i \cdot y_i}{\sum_{i=1}^N x_i} - \frac{\sum_{i=1}^N y_i}{N} \right)}{\left( \frac{\sum_{i=1}^N x_i^2}{\sum_{i=1}^N x_i} - \frac{\sum_{i=1}^N x_i}{N} \right)}, \quad (6)$$

$$b = \frac{\sum_{i=1}^N y_i}{N} - a \cdot \frac{\sum_{i=1}^N x_i}{N}. \quad (7)$$

Para terminar definimos el coeficiente de regresión  $R^2$  como una medida cuantitativa de lo acertado del ajuste. Así, un valor  $R^2$  cercano a 1 indica que el ajuste es bueno. La definición precisa es

$$R^2 = 1 - \frac{\chi^2}{\sum_{i=1}^N (y_i - \langle y_i \rangle)^2}, \quad (8)$$

donde

$$\langle y_i \rangle = \frac{1}{N} \sum_{i=1}^N y_i \quad (9)$$

es el promedio de los datos  $y_i$ .

## 1.2. Casos especiales

El método de regresión lineal se puede aplicar a otros modelos de interés físico. Revisaremos los tres casos más comunes.

1. **Caso semilog-y:** Se trata de modelos que tienen la forma

$$y^* = A \cdot e^{\alpha x}. \quad (10)$$

En este caso se toma logaritmo a ambos lados de la ecuación, resultando

$$\log(y^*) = \log(A) + \alpha \cdot x. \quad (11)$$

Se identifica entonces  $y = \log(y^*)$ ,  $a = \alpha$  y  $b = \log(A)$ , y se procede como se describió en la sección anterior. Al graficar los datos originales en un grafico semilog-y (eje  $y$  logarítmico, eje  $x$  lineal) se obtiene una línea recta. En Matlab un gráfico de este tipo se obtiene con la función *semilogy*( $x, y^*$ ) (Atención, esto es con los datos originales  $x$  e  $y^*$ ).

2. **Caso semilog-x:** Se trata de modelos de la forma

$$y = A \cdot \log(\alpha x^*). \quad (12)$$

Usando la propiedad de logaritmos de productos se obtiene

$$y = A \cdot \log(\alpha) + A \cdot \log(x^*). \quad (13)$$

Se identifica entonces  $x = \log(x^*)$ ,  $a = A$  y  $b = A \cdot \log(\alpha)$ , y se procede como en la sección anterior. En Matlab un gráfico de este tipo se obtiene con la función *semilogx*( $x^*, y$ ) (Atención, esto es con los datos originales  $x^*$  e  $y$ ).

3. **Ley de potencia:** Se trata de modelos de la forma

$$y^* = A \cdot (x^*)^\alpha. \quad (14)$$

Otra vez se toma logaritmo a ambos lados de esta ecuación y se obtiene

$$\log(y^*) = \log(A) + \alpha \cdot \log(x^*). \quad (15)$$

Se identifica en este caso  $y = \log(y^*)$ ,  $x = \log(x^*)$ ,  $a = \alpha$  y  $b = \log(A)$ . Al graficar los datos originales en un grafico log-log, se obtiene una línea recta. En Matlab un gráfico de este tipo se obtiene con la función *loglog*( $x^*, y^*$ ) (Atención, esto es con los datos originales  $x^*$  e  $y^*$ ). Este último caso es de interés para el objetivo de la sesión práctica asociada a esta guía.