Ajuste por mínimos cuadrados

En el ajuste por mínimos cuadrados, los datos de partida son un conjunto de puntos que forman la denominada "nube de puntos". El objetivo de este ajuste es obtener una función que marque la tendencia de la nube de puntos. La función puede ser de tipo exponencial, logarítmica, potencial, trigonométrica o polinómica. En el caso de ser polinómica, podemos elegir el grado del polinomio: primer grado (línea recta), segundo grado (parábola) o grado superior.

- Para comenzar el ejercicio, debemos elegir de entrada el tipo de función que vamos a emplear para el ajuste. Para ello, debemos analizar la nube de puntos y buscar el patrón de función que más se ajuste a ellos. **¡Ojo!:** En ajuste por mínimos cuadrados la función aproximadora no está obligada a pasar por todos los puntos del soporte, al contrario que en otros tipos de ajuste como la aproximación de Lagrange. Por tanto, el tipo de función o el grado del polinomio no están determinados por el número de puntos del soporte

  NOTA: si elegimos como función para el ajuste el polinomio de primer grado (línea recta) obtendremos la recta de regresión (recta azul de la Imagen 1).

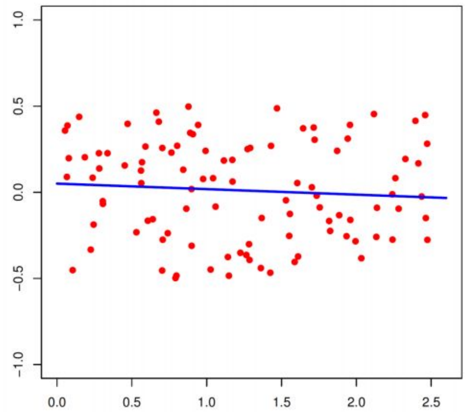


Imagen 1. Recta de regresión de la nube de puntos

- El ajuste obtenido con la función será mejor cuanto menor sea la distancia entre los puntos de la nube de puntos, y los puntos de la función representada. Dicha distancia se considera como un error en el ajuste.

- El **objetivo** es: minimizar la distancia entre la ordenada (valor en el eje y) de un punto de la nube de puntos y la ordenada del punto con la misma abscisa en la función representada. Es decir, para un mismo valor de x (por ejemplo, 1'5), se pretende minimizar la diferencia entre el valor que toma la x en la nube de puntos (por ejemplo, 0'5) y el valor que toma en la función representada (por ejemplo, 0'1).

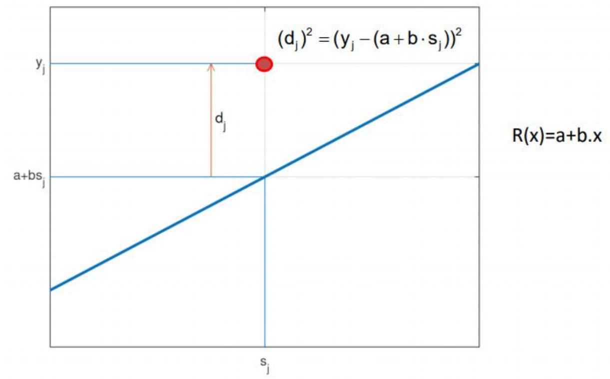


Imagen 2. Representación de la diferencia entre el punto

- ¿Por qué se toman mínimos cuadrados? Puesto que el objetivo es minimizar la distancia entre dos puntos, al elevar al cuadrado dicha distancia, la diferencia es mucho menor. Esto es debido a que las distancias suelen ser menores que la unidad. Tomando como ejemplo los datos anteriores, la diferencia entre los puntos es 0’4 (0’5-0’1=0’4). Sin embargo, si elevamos al cuadrado cada valor, la diferencia obtenida es 0’16 ($0,4^2$).
- Distinguimos entre 2 casos: el caso lineal y el no lineal. Sin embargo, cabe destacar que el caso lineal es una particularización del caso no lineal (cuando m=1) . Los estudiamos por separado para que sea más sencillo asentar los conceptos.

Ajuste por mínimos cuadrados. Caso lineal

En este caso, la función empleada es un polinomio de primer grado. La ecuación de la recta será: y=a+bx [donde *a* es la ordenada en el origen y *b* es la pendiente de la recta]. Buscamos determinar los coeficientes *a* y *b*

Con un solo punto la fórmula sería la siguiente:

$d_j = y_j - (a+b\cdot s_j)$

$(d_j)^2 = (y_j - (a+ b\cdot s_j))^2$

Pero como buscamos hacerlo para *n* puntos debemos aplicar un sumatorio

$$(d_j)^2 = f(a,b) = \sum_{j=1}^{n} ((y_j - (a + b\cdot s_j))^2$$

Queremos minimizar f (a, b). Para ello aplicamos el procedimiento de derivar e igualar a cero como en los problemas de optimización. La única diferencia es que la función depende de dos variables, por lo que derivamos parcialmente respecto a cada variable.

$$\frac{\partial f}{\partial a} = 2 \sum_{j=1}^{n} [(y_j - (a + b\cdot s_j)\cdot(-1)] = 0$$

$$\frac{\partial f}{\partial a} = 2 \sum_{j=1}^{n} [(y_j - (a + b\cdot s_j)\cdot(-s_j)] = 0$$

Si seguimos operando (multiplicamos por -1 en la ecuación 1 y por s_j en la 2 y despejamos):

1) $\sum_{j=1}^{n}[(y_j-(a+b\cdot s_j)\cdot(-1)]=0$ $\qquad \sum_{j=1}^{n} a+b\sum_{j=1}^{n} s_j=\sum_{j=1}^{n} y_j$

2) $\sum_{j=1}^{n}[(y_j-(a+b\cdot s_j)\cdot(-s_j)]=0$ →

$\sum_{j=1}^{n} a\cdot s_j+b\sum_{j=1}^{n}(s_j)^2=\sum_{j=1}^{n} y_j s_j$ →

NOTA: el sumatorio de una suma es la suma de sus sumatorios

¡Ojo!: No puedo sacar factor común de s_j y sacarlo del sumatorio, pues depende de la variable j

1) $na+b\sum_{j=1}^{n}[s_j]=\sum_{j=1}^{n} y_j$

2) $a\cdot\sum_{j=1}^{n} s_j+b\cdot\sum_{j=1}^{n}(s_j)^2=\sum_{j=1}^{n}(y_j\cdot s_j)$

NOTA: para resolver el sumatorio de una constante *a*, basta consumar la constante n veces, lo que equivale a *na*.

A continuación, con el fin de simplificar la ecuación, nombramos a los sumatorios con la siguiente nomenclatura:

$$SX\equiv\sum_{j=1}^{n} s_j;\quad SY\equiv\sum_{j=1}^{n} y_j;\quad SX2\equiv\sum_{j=1}^{n}(s_j)^2;\quad SSY\equiv\sum_{j=1}^{n}(y_j.s_j)$$

Con la nueva nomenclatura y las dos ecuaciones del apartado anterior, podemos construir el sistema:

$$\left\{\begin{array}{c} na+bSx=SY \\ aSX+bSX2=SSY \end{array}\right\}$$

Tenemos dos ecuaciones para obtener a, b mediante el método de CRAMER

$$\left\{\begin{array}{c} na+bSx=SY \\ aSX+bSX2=SSY \end{array}\right\} \Leftrightarrow A\cdot z=F; \quad A=\begin{pmatrix} n & Sx \\ Sx & SX2 \end{pmatrix}; z=\begin{pmatrix} a \\ b \end{pmatrix}; F=\begin{pmatrix} SY \\ SSY \end{pmatrix}$$

$$D=\begin{vmatrix} n & Sx \\ Sx & SX2 \end{vmatrix}=n\cdot SX2-(Sx)^2$$

$$a=\frac{\begin{vmatrix} SY & Sx \\ SSY & SX2 \end{vmatrix}}{D}=\frac{SY\cdot SX2-SX\cdot SSY}{D}=\frac{SY\cdot SX2-SX\cdot SSY}{n\cdot SX2-(Sx)^2}$$

$$b=\frac{\begin{vmatrix} n & SY \\ Sx & SSY \end{vmatrix}}{D}=\frac{n\cdot SSY-SX\cdot SY}{D}=\frac{n\cdot SSY-SX\cdot SY}{n\cdot SX2-(Sx)^2}$$

Teniendo en cuenta nuestra nomenclatura, podemos representar las ecuaciones de los parámetro a y b de la siguiente forma:

$$SX\equiv\sum_{j=1}^{n}s_j; \quad SY\equiv\sum_{j=1}^{n}y_j; \quad SX2\equiv\sum_{j=1}^{n}(s_j)^2; \quad SSY\equiv\sum_{j=1}^{n}(y_j.s_j)$$

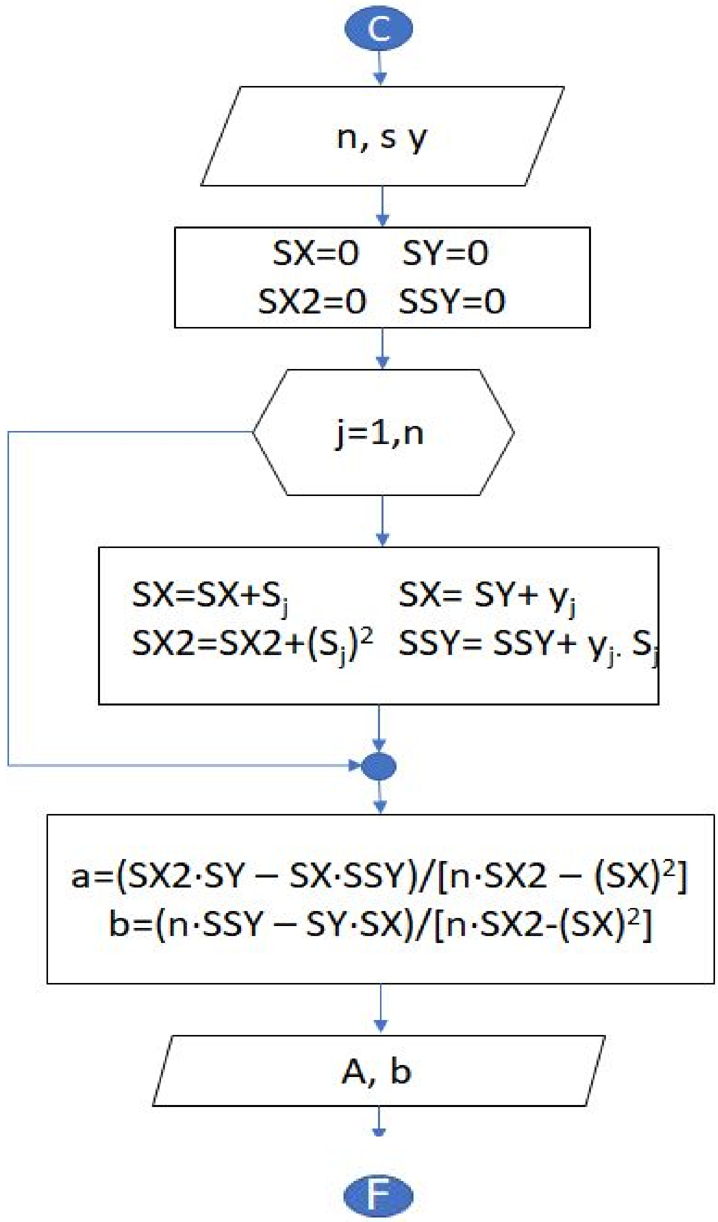
$$\Rightarrow\left\{\begin{array}{c} na+bSx=SY \\ aSX+bSX2=SSY \end{array}\right\}$$

$$a=\frac{\sum_{j=1}^{n}y_j\cdot\sum_{j=1}^{n}(s_j)^2-\sum_{j=1}^{n}s_j\cdot\sum_{j=1}^{n}(y_j.s_j)}{n\cdot\sum_{j=1}^{n}(s_j)^2-\left(\sum_{j=1}^{n}s_j\right)^2}$$

$$b=\frac{n\cdot\sum_{j=1}^{n}(y_j.s_j)-\sum_{j=1}^{n}s_j\cdot\sum_{j=1}^{n}y_j}{n\cdot\sum_{j=1}^{n}(s_j)^2-\left(\sum_{j=1}^{n}s_j\right)^2}$$

$$r(x)=a+bx$$

En la página siguiente proponemos un **organigrama para el cálculo de los coeficientes de la recta de regresión**, siendo '*s'* un vector de n componentes que contiene las abscisas de la nube de puntos e '*y'* otro vector de n componentes con las ordenadas de la nube de puntos:


C
n, s y
SX=0 SY=0
SX2=0 SSY=0
j=1,n
SX=SX+Sj SX= SY+ yj
SX2=SX2+(Sj)2 SSY= SSY+ yj. Sj
a=(SX2·SY – SX·SSY)/[n·SX2 – (SX)2]
b=(n·SSY – SY·SX)/[n·SX2-(SX)2]
A, b
F

Caso no lineal. Polinomio de grado M

Un polinomio de grado m, tendrá m+1 coeficientes: $P(x) = a_o + a_1 x + a_2 x^2 + ... + a_m x^m$. Por ejemplo, si un polinomio es de grado m=2, tendrá tres coeficientes: el término independiente, el término multiplicado por x y el término multiplicado por x^2.

Puesto que hay m+1 coeficientes, para poder ajustar la función harán falta m+1 ecuaciones, una por cada coeficiente del polinomio.

Así, en este caso, la fórmula de ajuste de mínimos cuadrados sería:

$$f(a_0, a_1, ..., a_m) = \sum_{j=1}^{n} d_j^{\,2} = \sum_{j=1}^{n} (y_j - P(s_j))^2$$

Si escribimos la ecuación anterior en función de cada coeficiente del polinomio obtenemos:

$$\sum_{j=1}^{n} [\, (y_j - \sum_{k=0}^{m} a_k \cdot x^k)^2]$$

donde *n* es el número de puntos en la nube y *m* es el grado del polinomio.

Si desarrollamos el sumatorio de los coeficientes del polinomio obtenemos:

$$\sum_{k=0}^{m} a_k \cdot x^k = a_o + a_1 s_j + a_2 s_j^{\,2} + a_3 s_j^{\,3} + ... + a_m s_j^{\,m}$$

Igual que en el caso lineal, para conseguir el mínimo error, debemos derivar la función respecto a cada uno de sus términos e igualarlo a cero. Por ello, se debe realizar la <u>derivada parcial</u> de la función respecto a cada término (se trata de un proceso análogo al que seguirías si resolvieras un problema de optimización para calcular el mínimo valor de una función).

$$\frac{\partial f}{\partial a_o} = 2 \cdot \sum_{j=1}^{n} [\, (y_j - (a_o + a_1 s_j + ... + a_m s_j^{\,m}))(-1)] = 0$$

$$\frac{\partial f}{\partial a_1} = 2 \cdot \sum_{j=1}^{n} [\, (y_j - (a_o + a_1 s_j + ... + a_m s_j^{\,m}))(-s_j)] = 0$$

$$\frac{\partial f}{\partial a_2} = 2 \cdot \sum_{j=1}^{n} [\, (y_j - (a_o + a_1 s_j + ... + a_m s_j^{\,m}))(-(s_j)^2)] = 0$$

$$\frac{\partial f}{\partial a_3} = 2 \cdot \sum_{j=1}^{n} [\, (y_j - (a_o + a_1 s_j + ... + a_m s_j^{\,m}))(-(s_j)^3)] = 0$$

…

$$\frac{\partial f}{\partial a_m} = 2 \cdot \sum_{j=1}^{n} [\, (y_j - (a_o + a_1 s_j + ... + a_m s_j^{\,m}))(-(s_j)^m)] = 0$$

A continuación, despejamos las ecuaciones:

$$\sum_{j=1}^{n} (a_o + a_1 s_j + ... + a_m s_j^{\,m}) = \sum_{j=1}^{n} y_j$$

$$\sum_{j=1}^{n} (a_o s_j + a_1 s_j^{\,2} + ... + a_m s_j^{\,m+1}) = \sum_{j=1}^{n} (s_j \cdot y_j)$$

$$\sum_{j=1}^{n} (a_o s_j^{\,2} + a_1 s_j^{\,3} + ... + a_m s_j^{\,m+2}) = \sum_{j=1}^{n} \ (s_j \cdot y_j^{\,2})$$

…

$$\sum_{j=1}^{n} (a_o s_j^{\,m} + a_1 s_j^{\,m+1} + ... + a_m s_j^{\,2m}) = \sum_{j=1}^{n} \ (s_j \cdot y_j^{\,m})$$

Por último, podemos representar el sistema de ecuaciones mediante una ecuación matricial para simplificar la expresión y facilitar la deducción de las fórmulas necesarias para resolver el algoritmo de ajuste por mínimos cuadrados en un caso no lineal. La ecuación matricial será de la forma $AX = b$ donde:

$$\mathbf{A} = \begin{pmatrix} \sum_{j=1}^{n} s_j^0 & \sum_{j=1}^{n} s_j & \sum_{j=1}^{n} s_j^2 & \sum_{j=1}^{n} s_j^3 & \cdots & \sum_{j=1}^{n} s_j^m \\ \sum_{j=1}^{n} s_j & \sum_{j=1}^{n} s_j^2 & \sum_{j=1}^{n} s_j^3 & \sum_{j=1}^{n} s_j^4 & \cdots & \sum_{j=1}^{n} s_j^{m+1} \\ \sum_{j=1}^{n} s_j^2 & \sum_{j=1}^{n} s_j^3 & \sum_{j=1}^{n} s_j^4 & \sum_{j=1}^{n} s_j^5 & \cdots & \sum_{j=1}^{n} s_j^{m+2} \\ \sum_{j=1}^{n} s_j^3 & \sum_{j=1}^{n} s_j^4 & \sum_{j=1}^{n} s_j^5 & \sum_{j=1}^{n} s_j^6 & \cdots & \sum_{j=1}^{n} s_j^{m+3} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum_{j=1}^{n} s_j^m & \sum_{j=1}^{n} s_j^{m+1} & \sum_{j=1}^{n} s_j^{m+2} & \sum_{j=1}^{n} s_j^{m+3} & \cdots & \sum_{j=1}^{n} s_j^{2m} \end{pmatrix}; \mathbf{X} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ \cdots \\ a_m \end{pmatrix}; \mathbf{b} = \begin{pmatrix} \sum_{j=1}^{n} y_j \\ \sum_{j=1}^{n} (s_j y_j) \\ \sum_{j=1}^{n} (s_j^2 y_j) \\ \sum_{j=1}^{n} (s_j^3 y_j) \\ \cdots \\ \sum_{j=1}^{n} (s_j^m y_j) \end{pmatrix}$$

Para obtener cada término de la matriz de los coeficientes podemos emplear la siguiente fórmula de sumatorio:

$A[i,n] = \sum_{j=1}^{n} s_j^{\,(i-1)+(k-1)} = \sum_{j=1}^{n} s_j^{\,i+k-2}$ (i=1,...,m+1) (k=1,...,m+1)

donde i representa el número de fila, k el número de columna y el sumatorio en j sirve para hacer el cálculo en los n puntos

Para obtener cada término de la matriz de los resultados (matriz b) podemos emplear el siguiente sumatorio:

$$b[i] = \sum_{j=1}^{n} y_j \cdot s_j^{\,(i-1)}$$

donde i representa el número de columnas y el sumatorio en j sirve para hacer el cálculo en los n puntos

En la siguiente página se muestra el **organigrama para obtener este sistema de ecuaciones a partir de la matriz.**