epsilon 66
2006

Una modificación del método de mínimos cuadrados ordinarios lineal

J.C. Cortés López
Departamento de Matemáticas Aplicadas
[Universidad Politécnica de Valencia]
G. Calbo Sanjuán
Departamento de Matemáticas. I.E.S. Els Évols. L'Alcúdia
[Valencia]

RESUMEN

En este trabajo se estudia una modificación del método de mínimos cuadrados ordinarios lineal, basada en la medición de las desviaciones en el mismo sentido que en geometría analítica plana se mide la distancia de un punto a una recta.

1. INTRODUCCIÓN Y PRELIMINARES

Una de las actividades estándar cuando se inicia el estudio de la geometría analítica en el plano, es el cálculo de la distancia D_i de un punto $P_i(x_i, y_i)$, a una recta dada en forma implícita, $Ax + By + C = 0$, la cual como es bien conocido está dada por

$$\frac{|Ax_i + By_i + C|}{\sqrt{A^2 + B^2}}. \quad \textbf{(1)}$$

En el caso particular, pero frecuente, en que la recta está expresada en la forma punto-pendiente, $r: y = bx + a$, dicha distancia vendrá dada por

$$D_i = \frac{|-bx_i + y_i - a|}{\sqrt{1 + b^2}}, \quad \textbf{(2)}$$

sin más que aplicar (1) para $A = -b$, $B = 1$ y $C = -a$, y . El valor (véase figura 1) es la longitud del segmento interceptado entre P_i y Q_i, siendo Q_i el punto de corte entre r y la recta perpendicular a r que pasa por el punto P_i (que es única). Sin embargo, en cursos posteriores (dentro de un nivel educativo universitario), cuando se estudia el método de mínimos cuadrados ordinarios (m.m.c.o.) lineal para calcular una recta de ajuste a una colección de puntos $P_i(x_i,y_i)$, $1 \leq i \leq n$, , se trata de minimizar (para garantizar la bondad del ajuste) las distancias (al cuadrado) en otro sentido: las desviaciones verticales d_i, aunque también existe la versión que trata de minimizar las distancias horizontales δ_i(véase figura 1). Una pregunta que surge de forma natural (sobre todo después de haber estudiado la geometría analítica elemental), es por qué no se considera en el m.m.c.o. minimizar las desviaciones en el sentido de involucrar las distancias D_i en lugar de las desviaciones verticales (horizontales) $d_i(\delta_i)$. La respuesta a esta cuestión es bien conocida: según el importante Teorema de Aproximación en espacios euclídeos de dimensión finita (véase, pág. 292 de Bru, 1998), la mejor elección de entre todas las posibles es la del m.m.c.o.

Conocedores de este resultado y buscando sobre todo enriquecer nuestra labor docente, nos hemos planteado reproducir la idea del m.m.c.o., pero introduciendo las distancias . En lo que sigue de artículo pretendemos comparar los conocidos resultados obtenidos para el m.m.c.o. lineal, que repasaremos en el siguiente apartado, con los que obtendremos aquí para lo que denominaremos en lo sucesivo el método de mínimos cuadrados modificado (m.m.c.m.).

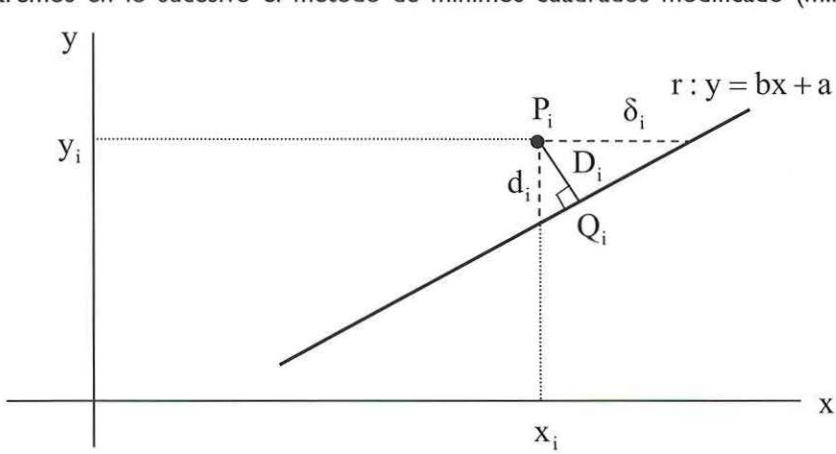

Figura 1. Método de mínimos cuadrados modificado (m.m.c.m.) lineal.

Para poder realizar un estudio comparativo del m.m.c.m. y el m.m.c.o., recordaremos las ideas y resultados básicos de este último, las cuales pueden ampliarse en (Bru, 1998) y en (pág. 164 de Barbolla, 2000). En el primer texto se trata esta temática desde una perspectiva algebraica, mientras que el enfoque del segundo está basado en el cálculo diferencial en varias variables.

Dada una colección de valores discretos o nube de puntos $\{(x_i,y_i)\}^n_{i=1}$, el objetivo del m.m.c.o. lineal es determinar una función lineal

$$s : y = bx + a, \quad \mathbf{(3)}$$

de modo que ajuste la nube de puntos en el siguiente sentido: minimizar el error cuadrático total E^2 dado por

$$E^2 = \sum_{i=1}^{n} d_i^2 \text{ siendo } d_i = y_i - (bx_i + a), \quad \mathbf{(4)}$$

o equivalentemente, minimizar el error cuadrático medio $\frac{E^2}{n}$ (véase figura 2). Obsérvese que (4) define el m.m.c.o. en su versión vertical, para la versión horizontal se procede de forma análoga.

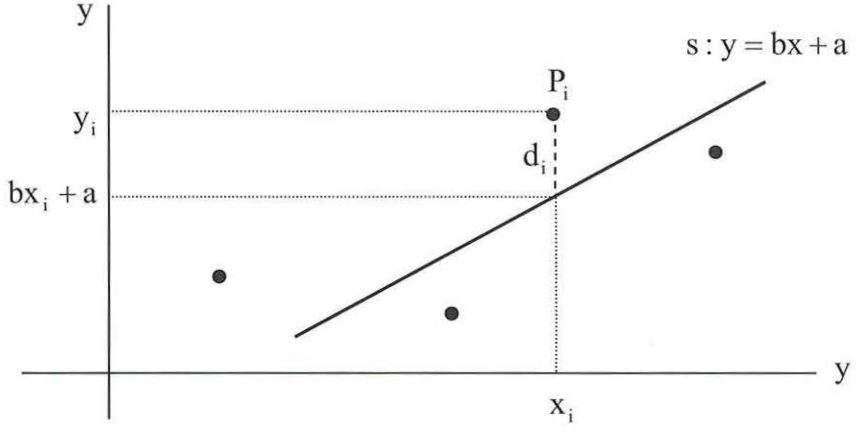

Figura 2. Método de mínimos cuadrados ordinario (m.m.c.o.) lineal.

Lo que más llama la atención del método es, a priori, la extraña forma o *sentido* de elegir el error a minimizar. Sin embargo, un análisis más detenido, nos desvela la más que adecuada elección del sentido de la aproximación o ajuste del m.m.c.o.:

- Si en lugar de minimizar en (4), $\sum_{i=1}^{n} d_i^2$, se tomase $\sum_{i=1}^{n} d_i$ (lo cual en principio puede parecer

  más natural), podría darse la no deseable, pero posible situación de que las desviaciones $d_1,\ldots,d_n$ respecto de la recta de ajuste fueran tales que se compensasen en signo y sucediese que, sin ser perfecto el ajuste, el error (ahora no cuadrático) total fuese nulo

$$\sum_{i=1}^{n} d_i = 0 .$$

- Una forma natural de evitar este problema, es considerar las desviaciones en valor absoluto, e intentar minimizar el error absoluto total

$$E = \sum_{i=1}^{n} |d_i| ,$$

  pero ello acarrearía un nuevo problema: sobre la función E, no podemos utilizar cómodamente métodos analíticos para minimizarla, pues al estar definida a través de un valor absoluto, existen problemas de derivabilidad.

Un enfoque que evita estos dos problemas es considerar la medición del error en el sentido del m.m.c.o. dado en (4): ahora no puede darse el posible problema de la compensación del signo de los errores al considerar las distancias al cuadrado, y dicha función siempre es derivable, por lo que las potentes técnicas de minimización basadas en el cálculo diferencial son ahora aplicables. Sin embargo, el precio que se paga por evitar los inconvenientes antes citados es que, en realidad no estamos midiendo el error o desviación del ajuste de forma natural, ya que, para un valor $(x_k, y_k) \in \{(x_i, y_i)\}^n_{i=1}$, la desviación vertical respecto de la recta de ajuste es $d_k = y_k - (bx_k + a)$, mientras que según el m.m.c.o. no se considera este valor sino d^2_k, i.e., aplicando el m.m.c.o. se deforman (por las razones antes indicadas) el verdadero valor de la desviación. De alguna manera con el m.m.c.o. premiamos y penalizamos las desviaciones $|d_k|$: tomando como patrón la distancia unidad, si la desviación es pequeña (grande), digamos $|d_k|<1$ ($|d_k|>1$) se considera d^2_k, que será en realidad un valor menor (mayor) que la desviación verdadera, premiando (penalizando) su proximidad (lejanía). En el caso en que $|d_k| = 1$ no se deforma la desviación.

Desde luego, los inconvenientes citados al principio y que resuelve en su *sentido* de aproximación el m.m.c.o. considerando las desviaciones al cuadrado (d_i^2), también quedan resueltos si se elige como función a minimizar la suma dada en (4), pero poniendo d_i^4, d_i^6, ...,d_i^{2m}, ... en lugar de d_i^2, aunque en estos casos se estarían deformando excesivamente las desviaciones reales, i.e., premiando y penalizando excesivamente las desviaciones de la recta de ajuste respecto de la nube de puntos $\{(x_i, y_i)\}^n_{i=1}$. Más aún, desde el punto de vista computacional al aumentar el exponente, los cálculos serían más engorrosos y además, como puede comprobarse fácilmente (véase (8), más adelante), la recta de ajuste no cumpliría la deseable propiedad (que sí se satisface en el caso del m.m.c.o.) de que pase por el centro de gravedad o baricentro de la nube de puntos.

Introduciendo la notación estadística para una colección de datos discretos bidimensionales $\{(x_i, y_i)\}_{i=1}^n$ dada en la tabla 1, llamando

$$m(a,b) = \sum_{i=1}^{n} [y_i - (bx_i + a)]^2 \quad \mathbf{(5)}$$

a la función a minimizar dada en (4) y aplicando el cálculo diferencial en dos variables, es sencillo calcular resolviendo el sistema de ecuaciones (que como es fácil comprobar es lineal)

$$\left.\begin{array}{l} \dfrac{\partial m}{\partial a} = 0 \\ \dfrac{\partial m}{\partial b} = 0 \end{array}\right\}, \quad \mathbf{(6)}$$

los valores $\tilde{a}$ y $\tilde{b}$ y que minimizan (5). Así, la recta dada en (3) que ajusta, en el sentido del m.m.c.o. la nube de puntos es $\{(x_i, y_i)\}_{i=1}^n$

$$s : y = \frac{\mu_{11}}{\sigma_x^2} x + (\overline{y} - \frac{\mu_{11}}{\sigma_x^2} \overline{x}) \quad \mathbf{(7)}$$

y se denomina *recta de regresión de Y sobre X*.

Obsérvese que como se cumple que

$$s: \overline{y} = \frac{\mu_{11}}{\sigma_x^2}\overline{x} + (\overline{y} - \frac{\mu_{11}}{\sigma_x^2}\overline{x}), \quad \textbf{(8)}$$

la recta de ajuste pasa por el centro de gravedad $(\overline{x}, \overline{y})$ de la nube de puntos, lo que da mayor solidez al m.m.c.o.

TABLA 1. Parámetros estadísticos muestrales bidimensionales.

MOMENTOS MUESTRALES DE UNA DISTRIBUCIÓN BIDIMENSIONAL $\{(x_i, y_i)\}_{i=1}^n$		
Momento respecto del origen de orden (r,s)	$\alpha_{rs} = \frac{1}{n}\sum_{i=1}^n x_i^r \cdot y_i^s$	
Momento central o respecto de la media de orden (r,s)	$\mu_{rs} = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^r \cdot (y_i - \overline{y})^s$	
CASOS PARTICULARES Y RELACIONES		
Media muestral	$\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i = \alpha_{10}$	$\overline{y} = \frac{1}{n}\sum_{i=1}^n y_i = \alpha_{01}$
Varianza muestral	$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2 = \mu_{20}$ $(\mu_{20} = \alpha_{20} - \overline{x}^2)$	$\sigma_y^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \overline{y})^2 = \mu_{02}$ $(\mu_{02} = \alpha_{02} - \overline{y}^2)$
Covarianza muestral	$\mu_{11} = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x}) \cdot (y_i - \overline{y}) = \alpha_{11} - \overline{x} \cdot \overline{y}$	

Cuando el ajuste por m.m.c.o. minimiza las distancias horizontales, la recta de ajuste, denominada *recta de regresión de* X *sobre* Y es

$$t: x = \frac{\mu_{11}}{\sigma_y^2}y + (\overline{x} - \frac{\mu_{11}}{\sigma_y^2}\overline{y})$$

que también satisface la propiedad de pasar por el baricentro $(\overline{x}, \overline{y})$ de la nube de puntos.

2. MODIFICACIÓN DEL M.M.C.O. LINEAL

Siguiendo la misma filosofía del m.m.c.o. lineal, pero trabajando con la modificación en el sentido especificado en el primer apartado, según (2), la función a minimizar para el m.m.c.m. será al sustituir el valor de despejado de (13) en (14) y después de simplificar se obtiene

$$(\alpha_{11}-\overline{x}\cdot\overline{y})b^2+(\alpha_{20}-\alpha_{02}+\overline{y}^2-\overline{x}^2)b+\overline{x}\cdot\overline{y}-\alpha_{11}=0\,. \qquad \textbf{(15)}$$

Supongamos que $\alpha_{11}\neq\overline{x}\cdot\overline{y}$ (o equivalentemente, véase tabla 1, $\mu_{11}\neq 0$), entonces dividiendo por $\alpha_{11}-\overline{x}\cdot\overline{y}$ la expresión (15), b se calcula a partir de la ecuación cuadrática

$$b^2+mb-1=0 \qquad \textbf{(16)}$$

siendo, por la tabla 1,

$$m=\frac{\alpha_{20}-\alpha_{02}+\overline{y}^2-\overline{x}^2}{\alpha_{11}-\overline{x}\cdot\overline{y}}=\frac{(\alpha_{20}-\overline{x}^2)-(\alpha_{02}-\overline{y}^2)}{\alpha_{11}-\overline{x}\cdot\overline{y}}=\frac{\sigma_x^2-\sigma_y^2}{\mu_{11}}. \qquad \textbf{(17)}$$

Como el discriminante de (16) es positivo: $\Delta_b=m^2+4>0$, , tendremos dos soluciones reales para $\hat{a}$ y $\hat{b}$:

$$\hat{a}_j=\overline{y}-\hat{b}_j\overline{x} \text{ siendo } \hat{b}_j=\begin{cases}-\dfrac{1}{2}m+\dfrac{\sqrt{m^2+4}}{2} & > \ 0\\ -\dfrac{1}{2}m-\dfrac{\sqrt{m^2+4}}{2} & < \ 0\end{cases},\ j=1,2. \qquad \textbf{(18)}$$

Por analogía al estudio que se hace habitualmente en el enfoque del m.m.c.o. a la pendiente $\hat{b}_j$ de la recta $y=\hat{a}_j+\hat{b}_jx$ le denominaremos *coeficiente de regresión de Y sobre X* y su valor indica la cuantía de variación del carácter dependiente por cada unidad de variación del carácter independiente X. Obsérvese que al igual que en el caso del m.m.c.o. el coeficiente de regresión "modificado" es adimensional por serlo tal y como se deduce de (17).

Por otra parte, por las relaciones de Cardano para (16) se tiene $\hat{b}_1\cdot\hat{b}_2=-1$, y como $\hat{b}_1$ y $\hat{b}_2$ son las respectivas pendientes de las rectas de ajuste que estábamos buscando, se deduce que el problema tiene dos candidatas a ser solución: dos rectas perpendiculares entre sí. La intuición geométrica (y la experiencia acumulada en el estudio del m.m.c.o.) nos indica que dados unos datos $\{(x_i,y_i)\}_{i=1}^n$ sólo una recta será la solución. Falta por tanto discriminar alguna candidata. Para hacerlo desde un marco teórico es necesario exigir la condición suficiente de segundo orden para mínimo descrita a través de la matriz hessiana:

$$\left.\frac{\partial^2 M}{\partial a^2}\right|_{(a,b)=(\hat{a}_j,\hat{b}_j)}>0 \quad , \quad \det\left(\left[\begin{matrix}\dfrac{\partial^2 M}{\partial a^2} & \dfrac{\partial^2 M}{\partial a\partial b}\\ \dfrac{\partial^2 M}{\partial b\partial a} & \dfrac{\partial^2 M}{\partial b^2}\end{matrix}\right]_{(a,b)=(\hat{a}_j,\hat{b}_j)}\right)>0\,.$$

Para la primera condición es sencillo comprobar que siempre se cumple, ya que:

$$\left.\frac{\partial^2 M}{\partial a^2}\right|_{(a,b)=(\hat{a}_j,\hat{b}_j)} = \frac{2n}{1+\hat{b}_j^2} > 0$$

pero, para la segunda, después de un considerable volumen de cálculo se llega a que

$$\det\left(\begin{bmatrix} \dfrac{\partial^2 M}{\partial a^2} & \dfrac{\partial^2 M}{\partial a \partial b} \\ \dfrac{\partial^2 M}{\partial b \partial a} & \dfrac{\partial^2 M}{\partial b^2} \end{bmatrix}_{(a,b)=(\hat{a}_j,\hat{b}_j)}\right) > 0 \Leftrightarrow \sigma_x^2 - \sigma_y^2 + 2\hat{b}_j \mu_{11} > 0. \quad \textbf{(19)}$$

Entonces, fijadas las observaciones $\{(x_i, y_i)\}_{i=1}^n$ distinguiremos varios casos:

Caso 1: $\mu_{11}>0$

Teniendo en cuenta (17), la condición (19) se puede escribir de la siguiente forma equivalente:

$$\sigma_x^2 - \sigma_y^2 + 2\hat{b}_j \mu_{11} > 0 \overset{\mu_{11}>0}{\Leftrightarrow} \frac{\sigma_x^2 - \sigma_y^2}{\mu_{11}} + 2\hat{b}_j > 0 \Leftrightarrow m + 2\hat{b}_j > 0 \Leftrightarrow \hat{b}_j > -\frac{1}{2}m \Leftrightarrow \hat{b}_j + \frac{1}{2}m > 0$$

de (18) se deduce que esta condición únicamente se cumple si allí elegimos $\hat{b}_1 > 0$. Resumiendo, si la covarianza de las observaciones es positiva, el coeficiente de regresión es positivo y debemos elegir la recta de regresión creciente o con pendiente positiva. Esto es exactamente lo mismo que sucede en el caso del m.m.c.o., como puede verse en (pág. 190 de Vargas, 1998).

Caso 2: si $\mu_{11}<0$

Razonando como antes se tiene:

$$\sigma_x^2 - \sigma_y^2 + 2\hat{b}_j \mu_{11} > 0 \overset{\mu_{11}<0}{\Leftrightarrow} \frac{\sigma_x^2 - \sigma_y^2}{\mu_{11}} + 2\hat{b}_j < 0 \Leftrightarrow m + 2\hat{b}_j < 0 \Leftrightarrow \hat{b}_j < -\frac{1}{2}m \Leftrightarrow \hat{b}_j + \frac{1}{2}m < 0$$

según (18), esta condición sólo se satisface si allí tomamos $\hat{b}_2 < 0$. Por tanto, si la covarianza de los datos es negativa, el coeficiente de regresión es negativo y se debe elegir la recta de regresión decreciente o con pendiente negativa, tal y como también sucede en el caso del m.m.c.o.

Caso 3: si $\mu_{11} = 0$

Obsérvese que ahora $\alpha_{11} = \overline{x} \cdot \overline{y}$. Suponiendo que (lo cual significa que la variabilidad observada de la variable $\sigma_x^2 = \mu_{20} > \mu_{02} = \sigma_y^2$ es mayor que la de la variable), entonces se satisfará la condición (19) de segundo orden para mínimo y como (15) ahora es

$$\underbrace{(\alpha_{11}-\overline{x}\cdot\overline{y})}_{=0}b^2+\underbrace{(\alpha_{20}-\alpha_{02}+\overline{y}^2-\overline{x}^2)}_{\sigma_x^2-\sigma_y^2>0}b+\underbrace{\overline{x}\cdot\overline{y}-\alpha_{11}}_{=0}=0$$

se tendrá: $\hat{b}=0$,, y de (13), $\hat{a}=\overline{y}$, por lo que la recta de ajuste es: $y=\overline{y}$. En palabras: cuando la covarianza es nula, el coeficiente de regresión es nulo y la recta de regresión de respecto de es paralela al eje de abscisas. Esto es exactamente lo mismo que sucede cuando se aplica el m.m.c.o. y la covarianza es nula. El caso μ_{11} = 0 y $\sigma_x^2=\mu_{20}<\mu_{02}=\sigma_y^2$, no podemos de momento abordarlo al cumplir, según (19) la condición de segundo orden de máximo (no de mínimo), pero más tarde cuando estudiemos la regresión de respecto de lo analizaremos. Por último señalar, que la posibilidad $\mu_{11}=0$ y $\sigma_x^2=\mu_{20}=\mu_{02}=\sigma_y^2$ y no es clasificable atendiendo al cálculo diferencial en varias variables, y requiere cuando se presente en la práctica un estudio particular.

Por analogía a como se procede al desarrollar el estudio del m.m.c.o. nos planteamos ahora cuál será la expresión de la recta de regresión de X respecto de Y mediante el m.m.c.m. Para su obtención, se deben buscar los coeficientes α y β de la recta

$$x=\beta y+\alpha$$

que minimiza la función

$$\Pi(\alpha,\beta)=\sum_{i=1}^{n}\Delta_i^2 \quad \text{siendo} \quad \Delta_i=\frac{|-x_i+\beta y_i+\alpha|}{\sqrt{1+\beta^2}},$$

tal y como es sencillo deducir al aplicar (2). Después de seguir un camino análogo al desarrollado anteriormente se llega a que el coeficiente β es solución de la ecuación

$$\beta^2-m\beta-1=0$$

(análoga a (16)) siendo el valor dado en (17), por lo que de nuevo existen dos soluciones candidatas

$$\hat{\alpha}_j=\overline{x}-\hat{\beta}_j\overline{y} \quad \text{siendo} \quad \hat{\beta}_j=\begin{cases}\dfrac{1}{2}m+\dfrac{\sqrt{m^2+4}}{2} & > \ 0\\ \dfrac{1}{2}m-\dfrac{\sqrt{m^2+4}}{2} & < \ 0\end{cases},\ j=1,2.$$

A la pendiente $1/\hat{\beta}$ de la recta x = βy + α le denominaremos *coeficiente de regresión de X sobre Y*, es adimensional, tiene una interpretación análoga a la dada para $\hat{b}_j$ y coinciden en el signo, como se deduce de la siguiente relación que se cumple:

$$\hat{b}_j = \frac{1}{\hat{\beta}_j}$$

como es sencillo comprobar, ya que, por ejemplo,

$$\hat{b}_1 \cdot \hat{\beta}_1 = \frac{-m + \sqrt{m^2+4}}{2} \cdot \frac{m + \sqrt{m^2+4}}{2} = 1 .$$

Por lo tanto, también en el caso de la regresión de X *sobre* Y obtenemos conclusiones similares a las anteriores: si la covarianza es positiva (negativa/nula), el coeficiente de regresión es positivo (negativo/nulo) y la recta de regresión es creciente (decreciente/paralela al eje de ordenadas). Realizando un desarrollo análogo al realizado para la regresión de X sobre Y, pero para la regresión de X sobre Y, se llegan a las condiciones de segundo orden:

$$\left.\frac{\partial^2 \Pi}{\partial a^2}\right|_{(a,b)=(\hat{a}_j,\hat{b}_j)} = \frac{2n}{1+\hat{b}_j^2} > 0 \quad ; \quad \det\left(\left[\begin{array}{cc} \frac{\partial^2 \Pi}{\partial a^2} & \frac{\partial^2 \Pi}{\partial a \partial b} \\ \frac{\partial^2 \Pi}{\partial b \partial a} & \frac{\partial^2 \Pi}{\partial b^2} \end{array} \right]_{(a,b)=(\hat{a}_j,\hat{b}_j)} \right) > 0 \Leftrightarrow \sigma_y^2 - \sigma_x^2 + 2\hat{b}_j \mu_{11} > 0$$

por lo tanto si presenta el caso (antes aparcado) en el que $\mu_{11} = 0$ y $\sigma_x^2 = \mu_{20} < \mu_{02} = \sigma_y^2$, la recta de ajuste que se debe tomar para minimizar los errores en el sentido del m.m.c.m. es la de regresión de X sobre Y.

3. EL COEFICENTE DE CORRELACIÓN LINEAL MODIFICADO

La correlación hace referencia al grado de relación entre dos variables, y el problema que se plantea es la determinación de una medida que indique el grado de intensidad de la relación entre las variables. En otras palabras, se trata de definir un valor que nos proporcione una medida del grado de ajuste de la curva a la nube de observaciones. Por la razones que se explican en (pág. 192 de Vargas, 1998), un buen indicador es el *coeficiente de correlación general de Pearson*, dado por

$$R = \pm\sqrt{1 - \frac{s_{ry}^2}{\sigma_y^2}}, \quad \textbf{(20)}$$

(a R^2 se le denomina *coeficiente de determinación*) donde s_{ry}^2 es la *varianza residual* de la variable dependiente definida a través de las diferencias entre los valores observados (y_i) y los teóricos $(y_i^* = \hat{a}_j + \hat{b}_j x_i)$:

$$s_{ry}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i^*\right)^2 . \quad \textbf{(21)}$$

Cuando la regresión es lineal este coeficiente se conoce con el nombre de *coeficiente de correlación lineal de Pearson* y suele denotarse por r. En el caso del m.m.c.o. lineal se tiene que

$$r = \frac{\mu_{11}}{\sigma_x \sigma_y}.$$

El problema que ahora nos planteamos es la determinación del correspondiente coeficiente de correlación lineal para el m.m.c.m. Para ello, en primer lugar evaluaremos la varianza residual según (21), teniendo en cuenta (10):

$$\begin{aligned}
s_{ry}^2 &= \frac{1}{1+\hat{b}_j^2}\frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i^*\right)^2 = \frac{1}{1+\hat{b}_j^2}\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{a}_j - \hat{b}_j x_i\right)^2 = \left\{\hat{a}_j = \bar{y} - \hat{b}_j \bar{x}\right\} \\
&= \frac{1}{1+\hat{b}_j^2}\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y} + \hat{b}_j \bar{x} - \hat{b}_j x_i\right)^2 = \frac{1}{1+\hat{b}_j^2}\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y} - \hat{b}_j(x_i - \bar{x})\right)^2 \\
&= \frac{1}{1+\hat{b}_j^2}\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 + \hat{b}_j^2 \frac{1}{n}\sum_{i=1}^{n}(\bar{x} - x_i)^2 - 2\hat{b}_j \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right]
\end{aligned}$$

es decir,

$$s_{ry}^2 = \frac{1}{1+\hat{b}_j^2}\left[\sigma_y^2 + \hat{b}_j^2 \sigma_x^2 - 2\hat{b}_j \mu_{11}\right] = \frac{1}{1+\hat{b}_j^2}\left[\left(\sigma_y^2 - \hat{b}_j \mu_{11}\right) + \underbrace{\hat{b}_j^2 \sigma_x^2 - \hat{b}_j \mu_{11}}_{(*)}\right] \quad \textbf{(22)}$$

Para simplificar más esta expresión obsérvese que por ser $\hat{b}_j$ solución de (16) se cumple, teniendo en cuenta (17), que

$$\hat{b}_j^2 + \frac{\sigma_x^2 - \sigma_y^2}{\mu_{11}}\hat{b}_j - 1 = 0,$$

si $\mu_{11} \neq 0$ (después abordaremos el caso $\mu_{11} = 0$), esto se escribe como

$$\hat{b}_j^2 \mu_{11} + \sigma_x^2 \hat{b}_j - \sigma_y^2 \hat{b}_j - \mu_{11} = 0,$$

como $\hat{b}_j \neq 0$ para j=1,2, (véase (18)), podemos multiplicar por $\hat{b}_j$ y obtenemos

$$\hat{b}_j^2 \sigma_x^2 - \hat{b}_j \mu_{11} = \hat{b}_j^2 \sigma_y^2 - \hat{b}_j^3 \mu_{11} \quad \textbf{(23)}$$

Sustituyamos ahora (23) en la expresión de (22) señalada con un asterisco, y obtenemos la siguiente expresión de la varianza residual

$$s_{ry}^2 = \frac{1}{1+\hat{b}_j^2}\left[\left(\sigma_y^2 - \hat{b}_j\mu_{11}\right) + \hat{b}_j^2\sigma_y^2 - \hat{b}_j^3\mu_{11}\right] = \frac{1}{1+\hat{b}_j^2}\left[\left(1+\hat{b}_j^2\right)\sigma_y^2 - \hat{b}_j\mu_{11}\left(1+\hat{b}_j^2\right)\right]$$

y simplificando

$$s_{ry}^2 = \sigma_y^2 - \hat{b}_j\mu_{11}.$$

Por lo tanto aplicando (20), el coeficiente el coeficiente de correlación lineal del m.m.c.m.

es $$\hat{R} = \hat{r} = \pm\sqrt{1 - \frac{\sigma_y^2 - \hat{b}_j\mu_{11}}{\sigma_y^2}} = \pm\frac{\sqrt{\hat{b}_j\mu_{11}}}{\sigma_y}. \quad \textbf{(24)}$$

Obsérvese que como se vio anteriormente, el radicando de (24) siempre está bien definido porque $\hat{b}_j$ y μ_{11} tienen el mismo signo, luego $\hat{b}_j\mu_{11} > 0$.

Esta relación entre el coeficiente de regresión de Y sobre X y el coeficiente de determinación modificado $\hat{R}^2$ modificado:

$$\hat{R}^2 = \hat{r}^2 = \frac{\mu_{11}}{\sigma_y^2}\hat{b}_j$$

permite expresar la recta de regresión Y sobre X en función de $\hat{r}^2$:

$$y - \overline{y} = \hat{r}^2\frac{\sigma_y^2}{\mu_{11}}\left(x - \overline{x}\right).$$

Anteriormente nos ha convenido aislar el caso μ_{11}=0, pero su estudio es más sencillo, ya que, sabíamos que en ese contexto si $\sigma_x^2 > \sigma_y^2$: $\hat{b} = 0$ y $\hat{a} = \overline{y}$, por lo que, la varianza residual en este caso vale

$$s_{ry}^2 = \frac{1}{1+\hat{b}^2}\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{a}_j - \hat{b}_j x_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2 = \sigma_y^2$$

y por tanto, sustituyendo en (20), R = 0.

Aunque un estudio análogo puede realizarse para la recta de regresión de sobre , por motivos de espacio y de su similitud con el desarrollo ya plasmado en estas páginas, no lo explicitaremos aquí.

4. APLICACIÓN

En este apartado, ilustraremos los resultados obtenidos para el m.m.c.m. aplicándolos al ajuste de una colección de datos, y comparando el error cuadrático que se obtiene con el correspondiente a la aplicación del m.m.c.o.

Considérese la siguiente distribución muestral bidimensional (véase tabla 2) recogida al valorar la compra de tierra para cultivo según diferentes índices de inflación

TABLA 2. Ejemplo de distribución muestral bidimensional.

x	3.9	4.5	5.2	6	5.4
y	2.1	2.8	2.4	3.1	2.8

El ajuste de los datos mediante un modelo lineal a través del m.m.c.o. conduce a la recta (véase pág. 304, Bru 1998)

$$s: \ y = 0.7881 + 0.3704x \ . \quad \textbf{(25)}$$

Para realizar el ajuste mediante el m.m.c.m. necesitamos previamente calcular los siguientes parámetros estadísticos:

$$\begin{array}{ll} \overline{x} = 5.0000 & \overline{y} = 2.6400 \\ \sigma_x^2 = 0.5320 & \sigma_y^2 = 0.1224 \\ \alpha_{11} = 13.3980 & \mu_{11} = 0.1980 \end{array}$$

entonces aplicando (17)

$$m = \frac{\sigma_x^2 - \sigma_y^2}{\mu_{11}} = 2.0686 \, .$$

Por otra parte, como $\mu_{11}>0$ de los dos valores posibles de $\hat{b}$ que proporciona (18), nos quedamos con el positivo

$$\hat{b} = \frac{-2.0686 + \sqrt{2.0686^2 + 4}}{2} = 0.4044 \, .$$

(esto es intuitivo: como la nube de puntos (véase figura 3) marca claramente una tendencia creciente o de pendiente positiva, la pendiente de la recta de regresión del m.m.c.m. debe ser positiva). Aplicando (13), $\hat{a}$ = 2.6400 - 0.4044 . 5 = 0.6180. Por lo tanto la recta de ajuste buscada por el m.m.c.m. es

$$r: \ y = 0.6180 + 0.4044x \ . \quad \textbf{(26)}$$

En la figura 3, se muestra el ajuste de la distribución bidimensional de la tabla 2 por ambos métodos.

Para terminar calculamos los errores cuadráticos. Para ello y con la notación introducida, calculamos el error cuadrático respecto de ambas rectas y , medidos en el sentido de los m.m.c.o. y m.m.c.m.: E_r^2, $\hat{E}_r^2$, E_s^2 y $\hat{E}_s^2$.

$$
\begin{aligned}
E_s^2 &= (2.1-2.2327)^2+(2.8-2.4549)^2+ \\
&+ (2.4-2.7142)^2+(3.1-3.0105)^2+(2.8-2.7883)^2 = 0.2436
\end{aligned}
$$

$$
\begin{aligned}
E_r^2 &= (2.1-2.1952)^2+(2.8-2.4378)^2+ \\
&+ (2.4-2.7209)^2+(3.1-3.0444)^2+(2.8-2.8018)^2 = 0.2463
\end{aligned}
$$

$$
\begin{aligned}
\widetilde{E}_s^2 &= \frac{(-0.3704\cdot 3.9+2.1-0.7881)^2}{1^2+0.3704^2} + \frac{(-0.3704\cdot 4.5+2.8-0.7881}{1^2+0.3704^2} \\
&+ \frac{(-0.3704\cdot 5.2+2.4-0.7881)^2}{1^2+0.3704^2} + \frac{(-0.3704\cdot 6+3.1-0.7881)^2}{1^2+0.3704^2} \\
&+ \frac{(-0.3704\cdot 5.4+4.8-0.7881)^2}{1^2+0.3704^2} = 3.7728
\end{aligned}
$$

$$
\begin{aligned}
\widetilde{E}_r^2 &= \frac{(-0.4044\cdot 3.9+2.1-0.6180)^2}{1^2+0.4044^2} + \frac{(-0.4044\cdot 4.5+2.8-0.6180}{1^2+0.4044^2} \\
&+ \frac{(-0.4044\cdot 5.2+2.4-0.6180)^2}{1^2+0.4044^2} + \frac{(-0.4044\cdot 6+3.1-0.6180)}{1^2+0.4044^2} \\
&+ \frac{(-0.4044\cdot 5.4+4.8-0.6180)^2}{1^2+0.4044^2} = 3.9434
\end{aligned}
$$

Obsérvese que, como adelantamos en el primer apartado la recta de ajuste por m.m.c.o. obtiene menor error cuadrático tanto si se miden las distancias en el sentido ordinario, es decir, $E_s^2 < E_r^2$, como si se hace en el sentido de la geometría analítica: $\widetilde{E}_s^2 < \widetilde{E}_r^2$, , aunque el ajuste por el m.m.c.m. vemos que es prácticamente equivalente al obtenido por el m.m.c.o.

Para terminar, señalar que la generalización del m.m.c.m. a más de dos variables (regresión multivariante) puede realizarse de una forma análoga a la aquí realizada pero requiere el uso de los correspondientes resultados de cálculo diferencial en varias variables.

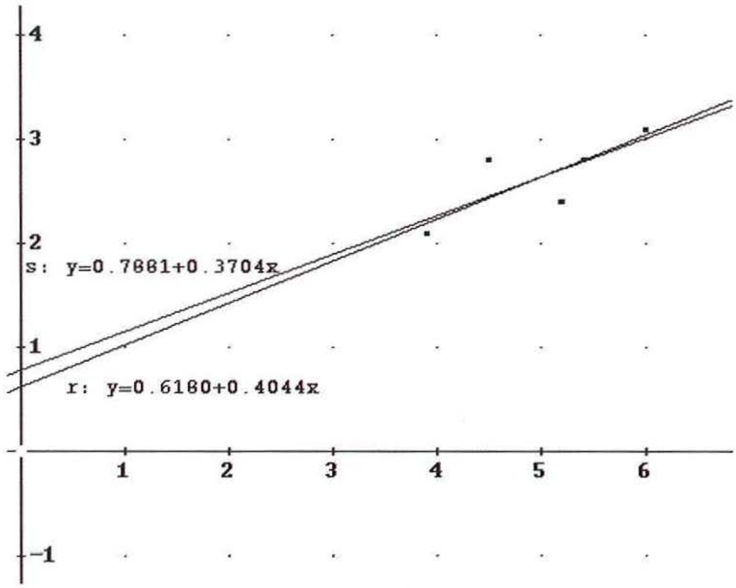

Figura 3. Ajuste de la tabla 2 por el m.m.c.m. (r) y por el m.m.c.o. (s).

5. CONCLUSIONES

En el trabajo se estudia una variación del método de mínimos cuadrados ordinario lineal basada en la medición de las desviaciones en otro sentido: el mismo en que en geometría métrica plana se mide la distancia de un punto a una recta. A sabiendas de que el nuevo método acarreará un error cuadrático mayor que el que supone el m.m.c.o., el estudio del método es interesante y muy natural en su planteamiento para los alumnos de un primer curso universitario que proceden de estudiar problemas geométricos en el bachillerato. En este sentido, creemos que el trabajo puede tener interés docente, para ser aprovechado en el aula a través de un estudio comparativo con el m.m.c.o.

BIBLIOGRAFÍA

BARBOLLÁ R., CERDÁ E. y SANZ P. (2000). *Optimización,* Madrid, Ed. Prentice Hall.

BRU R., CLIMENT J.J., MAS J. y URBANO A. (1998). *Álgebra Lineal,* Valencia, Ed. Universidad Politécnica de Valencia.

VARGAS A. (1998). *Estadística Descriptiva e Inferencial,* 1ª reimpresión, Cuenca, Ed. Universidad de Castilla La Mancha, Nº 8 Colecc. Ciencia y Técnica.