SESIÓN DE TRABAJO DEL FORO PERMANENTE DE PROFESORES DE CARRERA DE LA DIVISIÓN DE CIENCIAS BÁSICAS

"Mínimos Cuadrados"

M. en I. Leda Speziale San Vicente

17-09-2008

Buenas tardes, vamos a dar inicio a la sesión de hoy, en esta ocasión tenemos a la maestra Leda Speziale San Vicente y el título de la ponencia es "Mínimos Cuadrados" entonces, maestra le cedo la palabra.

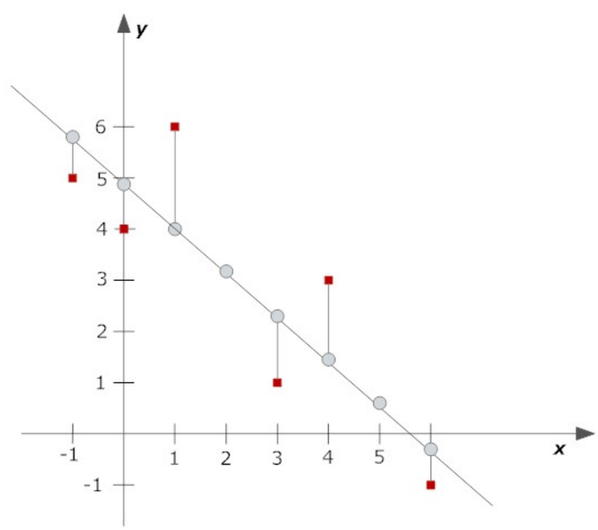
Muchas gracias, buenas tardes.

"El Método de Mínimos Cuadrados" surgió ya hace mucho tiempo, la presentación actual está basada en conceptos de Álgebra Lineal que es una forma moderna de ver las cosas, más elegante, más versátil y que tiene muchas aplicaciones.

Hablaremos del nombre y en qué consiste.

Tenemos un conjunto de parejas de valores, expresado en una tabla, que relacionan a dos variables, la *x* y la *y*, que se han obtenido por medio de alguna experiencia y deseamos saber qué relación matemática pueden tener, para ello se grafican y obtenemos que, por el aspecto de la gráfica o por conocimiento del experimento al que se refieren, podemos suponer que la relación entre los valores de estas variables es lineal. Por lo tanto, queremos ajustar una línea recta, es decir, necesitamos conocer las constantes que nos definen la ecuación de la línea recta: la pendiente y la ordenada al origen. Si tenemos el conjunto de seis parejas de valores:

x	y
−1	5
0	4
1	6
3	1
4	3
6	−1



seis valores de la *x* y seis valores de la *y*, podemos plantear un sistema de seis ecuaciones con dos incógnitas (la *m* y la *b*), pendiente y ordenada al origen de la ecuación de la recta $y = mx + b$, lógicamente por ser algo surgido de la experiencia, este sistema resulta incompatible. ¿Cuándo resulta compatible?, cuando existe una recta que pase por todos los puntos, y eso generalmente no sucede, porque son valores empíricos, entonces se trata de buscar la línea recta que se acerque más a esos valores. El método se llamó desde un principio de Mínimos Cuadrados, pues existe una diferencia entre el valor observado y el valor teórico correspondiente a la línea recta que buscamos: la diferencia entre la *y* observada y la *y* que correspondería a la recta; a estas diferencias se les elevó al cuadrado y, realmente lo que se minimiza es "la suma de los cuadrados" de dichas diferencias, que en un principio se llamaban errores, entonces, para que los errores fueran mínimos, se consideraba la suma de los cuadrados de esos errores, ¿por qué los cuadrados?, para que no influyera el que estuvieran de forma negativa o de forma positiva, para eliminar ese factor del signo y la suma de los cuadrados se minimizaba. Ahora, con conceptos de Álgebra Lineal el procedimiento se ha transformado, en mi opinión, en una forma más elegante y además un poquito más versátil, como les decía al principio.

Si expresamos el sistema en forma matricial: $A\overline{u} = \overline{y}$, la matriz de coeficientes multiplicada por el vector de incógnitas equivale a multiplicar la matriz de coeficientes por el vector $\begin{bmatrix} m \\ b \end{bmatrix}$ y debido a la multiplicación de matrices, eso resulta equivalente a tener una columna por la primera de las incógnitas más la otra columna por la otra de las incógnitas. Cada columna considerada como un vector de R^6 por una constante, es decir, una combinación lineal de las columnas.

El sistema tiene solución siempre y cuando el vector formado por los términos independientes pertenece al espacio generado por las columnas de la matriz de coeficientes; a ese vector lo llamamos $\overline{y_1}$, lo que queremos es encontrar este vector solución de este sistema de seis ecuaciones lineales, que se acerque más al vector $\overline{y}$ observado. Para ello vamos a utilizar conceptos de Álgebra Lineal. Como vamos a trabajar con producto interno y tenemos aquí eneadas de números reales (porque son valores observados) que pertenecen el espacio R^6, vamos a trabajar con el producto interno que se conoce como el producto interno usual en R^n que es el producto escalar o producto punto, ¿y por qué vamos a usar eso?, porque la norma de un vector con el producto punto es la raíz cuadrada de la suma de los cuadrados de sus componentes. También vamos a manejar lo que en Álgebra Lineal se conoce como el Teorema de Proyección, que dice: existe un vector $\overline{w}$ que pertenece a un subespacio de un espacio *V*, tal que, para algún vector que pertenezca a *V*, el $\overline{w}$ es el que tiene la distancia mínima, es decir, del que la norma de la diferencia (que corresponde precisamente a la distancia), es menor o igual que la distancia que corresponde a cualquier otro vector de ese subespacio. A éste, que es el vector más cercano a un vector del espacio *V* en el subespacio que estamos llamando aquí *W*, se le acostumbra llamar la proyección del vector de *V* sobre el subespacio *W*, por esto se llama Teorema de Proyección, que por supuesto no demostramos aquí.

Tenemos también otro teorema que dice que para todo vector de un cierto espacio *V,* existe un vector $\overline{w}$ que pertenece a *W* y otro, que aquí estoy llamando $\overline{g}$, que pertenece al ortogonal de este subespacio *W*, tal que el vector cualquiera $\overline{v}$ del espacio, siempre se puede expresar como una suma de $\overline{w}$ más $\overline{g}$ que pertenece al complemento ortogonal del espacio; este $\overline{w}$ es precisamente la proyección del vector $\overline{v}$ sobre el subespacio *W*.

Utilizando estos tres conceptos fundamentales del Álgebra Lineal: producto interno el usual en R^n, producto escalar o producto punto; teorema de proyección y este último teorema, tenemos otro enfoque, otra forma de manejar el Método de Mínimos Cuadrados, como un método moderno; el Álgebra Lineal es una álgebra que no es de hace cientos de años, sino más bien es una álgebra moderna, y tiene ventajas. Entonces, considerando el ejemplo:

$$\begin{aligned} -m + b &= 5 \\ 0m + b &= 4 \\ m + b &= 6 \\ 3m + b &= 1 \\ 4m + b &= 3 \\ 6m + b &= -1 \end{aligned} \qquad A\overline{u} = \overline{y} \qquad \overline{u} = \begin{bmatrix} m \\ b \end{bmatrix} \qquad A = \begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 3 & 1 \\ 4 & 1 \\ 6 & 1 \end{bmatrix}$$

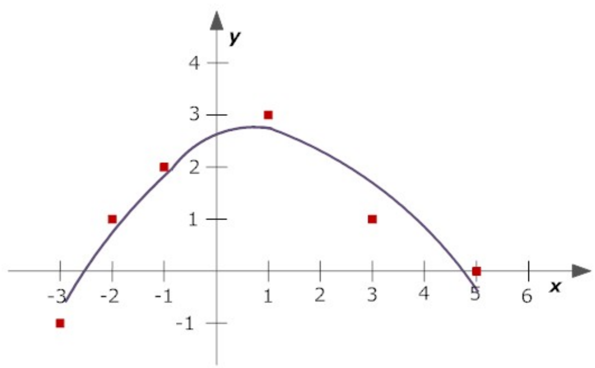
el vector $\overline{y}$ de valores empíricos pertenece a R^6 (en este ejemplo se tiene un conjunto de seis parejas de valores, pero esto utilizando la computadora, se puede llevar a un conjunto más grande; además, manejar las matrices con la computadora es mucho más fácil; aquí esta manejado, digamos a mano, para que se vea qué es lo que sucede), y por el teorema último que vimos, existe un vector $\overline{y_1}$ que pertenece al espacio generado por las columnas de la matriz *A* (que corresponde al vector de términos independientes del sistema de ecuaciones que sí tiene solución, que es compatible) y un $\overline{g}$ que pertenece al complemento ortogonal del espacio al que pertenece $\overline{y_1}$, por lo tanto, tenemos que $\overline{y} = \overline{y_1} + \overline{g}$. Ahora, si $\overline{g}$ es un vector del complemento ortogonal debe ser ortogonal tanto a la primera columna como a la segunda columna de la matriz *A*, así tenemos que, el producto de la primera columna de *A* por el vector $\overline{g}$ debe ser igual a cero y también el producto con la segunda columna; esto nos conduce a un sistema de dos ecuaciones con seis incógnitas, ¿Quiénes son las incógnitas?, pues las componentes de $\overline{g}$. Ahora, este sistema se puede expresar de la manera siguiente:

$$\begin{aligned} \overline{C_{1A}} \cdot \overline{g} &= 0 \\ \overline{C_{2A}} \cdot \overline{g} &= 0 \end{aligned} \quad \Rightarrow \quad \begin{aligned} a_{11}g_1 + a_{21}g_2 + a_{31}g_3 + a_{41}g_4 + a_{51}g_5 + a_{61}g_6 &= 0 \\ a_{12}g_1 + a_{22}g_2 + a_{32}g_3 + a_{42}g_4 + a_{52}g_5 + a_{62}g_6 &= 0 \end{aligned} \quad ; \quad A^T\overline{g} = \overline{0}$$

donde la matriz de coeficientes es la transpuesta de *A.* Por otro lado, si despejamos a $\overline{g}$ de $\overline{y} = \overline{y_1} + \overline{g}$ resulta $A^T\left(\overline{y} - \overline{y_1}\right) = \overline{0}$ donde $\overline{y_1}$ es el vector que queremos que sea exactamente igual a $A\overline{u}$ y llegamos a: $A^T\left(\overline{y} - A\overline{u}\right) = \overline{0}$ que nos conduce a las ecuaciones que generalmente se llaman "las ecuaciones normales del Método de Mínimos Cuadrados": $A^TA\overline{u} = A^T\overline{y}$ donde $\overline{u}$ es el vector de incógnitas. Entonces para nuestro ejemplo llegamos a los valores de *m* y de *b* que son los que corresponden precisamente a la recta que mejor se ajusta por el método de mínimos cuadrados al conjunto de valores, esto es en términos generales la teoría del método de mínimos cuadrados pero con conceptos de Álgebra Lineal. Como dije se puede hacer más versátil, se puede aplicar a más casos y traemos aquí algunos ejemplos.

Por ejemplo, tenemos un conjunto de valores observados y esos valores son precisamente estos puntos que no corresponden a una línea recta, entonces puede ser que por el aspecto de la gráfica de los valores, o bien porque sabemos a qué experimento observado se refieren y tenemos idea de que la relación es una relación correspondiente a un polinomio de segundo grado, vamos a tratar de ajustar un polinomio de segundo grado: $y_1 = ax^2 + bx + c$

x	y
−3	−1
−2	1
−1	2
1	3
3	1
5	0



ahora las incógnitas ya no son *pendiente* y *ordenada al origen* sino los tres coeficientes, ya que no es una línea recta sino un polinomio de segundo grado y esto es mucho fácil manejarlo con los conceptos de Álgebra Lineal. Aquí la matriz de coeficientes tiene como primera columna los cuadrados de los valores observados, la segunda columna contiene los valores correspondientes al coeficiente *b,* o sea*,* los valores de *x*, y la tercera los del término independiente. Así utilizando las ecuaciones normales que ya vimos tenemos:

$$A\bar{u} = \bar{y}$$

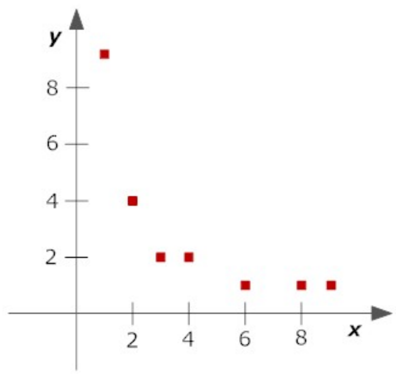
$$\bar{u} = \begin{bmatrix} m \\ b \end{bmatrix}$$

$$A = \begin{bmatrix} 9 & -3 & 1 \\ 4 & -2 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \\ 9 & 3 & 1 \\ 25 & 5 & 1 \end{bmatrix}$$

$$A^T A\bar{u} = A^T \bar{y}$$

La *A* es de 6 por 3, entonces la transpuesta es de 3 por 6 y el producto de la transpuesta por *A* resulta de 3 por 3 y queda un sistema de tres ecuaciones con tres incógnitas, y el vector de términos independientes es el producto de A^T por $\bar{y}$ que es de 3 por 1. Así resulta mucho más fácil que estar buscando la diferencia, que es lo que antiguamente se llamaban errores, en términos de *a, b,* y *c* y luego derivar esto parcialmente para poder encontrar el mínimo de la suma de los cuadrados de los errores; y con Álgebra Lineal todo es inmediato, debido a que encontrando la matriz de coeficientes, la transponemos, hacemos el producto y usamos la ecuación normal que algunos llaman también *Ordinaria.*

Como decíamos, para aplicar este procedimiento necesitamos que la función sea un polinomio para tener un sistema de ecuaciones lineales. Sin embargo, en algunas ocasiones se tiene un conjunto de puntos como en este tercer ejemplo cuya relación se asemeja más a una exponencial como $y_1 = ae^{bx}$:



Para trabajar con los conceptos de Álgebra Lineal es necesario ingeniarse y convertir esta relación exponencial en una polinomial. ¿Qué se propone?, pues tomar el logaritmo natural tanto del primer miembro como del segundo, con lo que queda $Ly_1 = La + bx$. Necesitamos encontrar el vector $\overline{Ly_1}$ más próximo al vector $\overline{Ly}$ que se obtiene con los valores observados.

x	y
1	9
2	4
3	2
4	2
6	1
8	1
9	1

Como puede verse la ecuación $Ly_1 = La + bx$ equivale a la ecuación de una recta con pendiente igual a b y ordenada al origen igual a La. Aplicando el método visto se tiene

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 6 \\ 1 & 8 \\ 1 & 9 \end{bmatrix} \quad \text{y} \quad \overline{u} = \begin{bmatrix} La \\ b \end{bmatrix}$$

en el sistema

x	y	Ly
1	9	2.197
2	4	1.386
3	2	0.693
4	2	0.693
6	1	0
8	1	0
9	1	0

$A^T A\overline{u} = A^T \overline{Ly}$, con lo que

$$\begin{bmatrix} 7 & 33 \\ 33 & 24 \end{bmatrix} \begin{bmatrix} La \\ b \end{bmatrix} \approx \begin{bmatrix} 4.97 \\ 9.81 \end{bmatrix} \text{ de donde } \begin{cases} La \approx 1.868 & \Rightarrow \quad a \approx 6.48 \\ b \approx -0.246 \end{cases} \quad \therefore \quad y_1 \approx 6.48e^{-0.246x}.$$

Con un poco de ingenio el método también lo podemos aplicar cuando tenemos valores observados en coordenadas polares, por ejemplo, si se tiene una serie de coordenadas r y θ y queremos conocer los valores de a y b de la ecuación de la cónica más cercana a los puntos correspondientes a los valores observados. La ecuación es $r = \dfrac{a}{1 - b\cos\theta}$, que se puede expresar como $r = br\cos\theta + a$ y que corresponde a la de una recta con pendiente b y cuya ordenada al origen es a, con variable dependiente r y variable independiente $r\cos\theta$. Si tenemos los siguientes valores:

θ	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{7}{12}\pi$	$\frac{2}{3}\pi$
r	4	4	3.4	3.0	2.6	2.6
$r\cos\theta$	3.46	2.83	1.70	0	-0.67	-1.30

$$A = \begin{bmatrix} 3.46 & 1 \\ 2.83 & 1 \\ 1.70 & 1 \\ 0 & 1 \\ -0.67 & 1 \\ -1.30 & 1 \end{bmatrix} \quad \text{en} \quad A^T A\bar{u} = A^T \bar{r}$$

$$A^T A \approx \begin{bmatrix} 25 & 6 \\ 6 & 6 \end{bmatrix} \; ; \; A^T \bar{r} \approx \begin{bmatrix} 25.8 \\ 19.6 \end{bmatrix} \quad \therefore \quad \bar{u} = \begin{bmatrix} b \\ a \end{bmatrix} \approx \begin{bmatrix} 0.32 \\ 2.94 \end{bmatrix}$$

¿Qué cónica es?, depende del valor de b, puede ser desde una circunferencia si es cero, pues queda el radio igual a una constante; si es 1 es parábola; si b está entre 0 y 1 es elipse; y si es mayor que 1 es hipérbola. Como en nuestro caso es 0.32, entonces, la curva más cercana a los puntos empíricos es una elipse, de la que podemos encontrar su centro, su semieje mayor y su semieje menor al convertir su ecuación polar a cartesiana.

En términos generales esto es todo, espero que les haya sido de interés, sobre todo lo que es el enfoque de Álgebra Lineal, que yo siento que es mucho más elegante, mucho más bonito y más versátil que como se manejaba antes.

Yo desde que empecé a dar Probabilidad, ayer, en 1968, algo así, conocí este método de *El Ajuste de Curvas por Mínimos Cuadrados*, supongo, me corregirán los que están aquí presentes, que en física se sigue usando dicho método y creo que es conveniente darle el enfoque de Álgebra Lineal.