Comunicaciones en Estadística
Agosto 2016, Vol. 9, No. 2, pp. 223–237

Modalidad didáctica histórico-evolutiva para la regresión[1]

Historical-Evolutionary teaching mode for regression

Alfonso Sánchez H.[a]
asanchez@ut.edu.co

Silvia Sarzoza H.[b]
ssarzoza@upla.cl

José A. González C.[c]
jgonzalez@upla.cl

Resumen

El proceso de enseñanza y aprendizaje está sufriendo un cambio en su paradigma regulador, haciendo real y operativa la innovación pedagógica, que supone la transición desde un modelo entrado en la enseñanza hacia un modelo centrado en el aprendizaje del estudiante. Esta modificación invita a cuestionarse qué pretendemos que aprendan los estudiantes, cuáles son las modalidades y metodologías más adecuadas para que ello puedan adquirir estos aprendizajes y con qué criterios y procedimientos vamos a comprobar si los ha adquirido finalmente. Este artículo presenta una modalidad didáctica histórica-evolutiva para la regresión con base en métodos alternativos.

***Palabras clave*:** regresión, modalidad didáctica, modalidad de enseñanza, aprendizaje.

Abstract

The teaching and learning process is undergoing a change in its regulatory paradigm, making real and operational pedagogical innovation that marks the transition from a model focused on teaching to a model focused on student learning. This modification invites into question what we intend that students learn, what

[1]Sánchez, A., Sarzoza, S., González J. Modalidad didáctica histórico-evolutiva para la regresión. *Comunicaciones en Estadística*, **9**(2), 223-237.

[a]Universidad de Tolima, Grupo GELIMO, Departamento de Matemática y Estadística, Ibagué, Colombia

[b]Universidad de Playa Ancha, Facultad de Ciencias de la Educación, Centro de Estudios Avanzados, 2340000, Valparaíso, Chile.

[c]Universidad de Playa Ancha, Facultad de Ciencias Naturales y Exactas, Departamento de Matemática y Estadística, Laboratorio de Investigación y Experimentacíon de Saberes Matemáticos, Valparaíso, Chile.

are the modalities and most appropriate for the student to acquire these learnings and what criteria and methodologies are procedures will check if the student has finally acquired. This article brings to present a historical-evolutive teaching method for alternative methods based on regression.

***Keywords*:** regression, teaching method, teaching mode, learning.

1. Introducción

No basta con decir que el profesor debe cambiar el rol a la hora de planificar su docencia, sino que es necesario presentar modelos y pautas que le ayuden en este proceso. Precisamente porque se quiere evitar este riesgo, se entiende que es necesario promover estudios y modelos orientados a generar dentro del profesorado universitario una cultura favorable hacia el cambio de paradigma en los procesos de enseñanza, al tiempo que debe acompañarse de procedimientos instrumentales que faciliten abordar esta tarea. Y esto es lo que precisamente se propone en este artículo, para lo cual se establece un conjunto de orientaciones técnicas que permitan al profesorado distribuir su actividad docente en diversas modalidades y metodologías de enseñanza y aprendizaje (de Miguel 2005).

Resulta prioritario impulsar estudios que faciliten estrategias a los profesores universitarios, para promover la adquisición del aprendizaje, a partir de la interacción de factores, tanto personales como contextuales. Es decir, de la relación directa entre el modelo de enseñanza, el modelo de evaluación y la forma en que el estudiante aborda el aprendizaje desde su propia perspectiva (Entwistle & Ramsden 2015, Ramsden 2003, Biggs & Biggs 2005, Buendía & Olmedo 2002).

La forma en que el estudiante enfrenta las situaciones de aprendizaje se ve influenciada por la percepción que tiene respecto al medio y su relación con la enseñanza, la que se define a partir de las concepciones de los profesores respecto a esta y la implicancia en el aprendizaje de sus estudiantes. Para que los estudiantes consigan los resultados deseados de una manera razonablemente eficaz, la tarea fundamental del profesor consiste en lograr que sus alumnos realicen las actividades de aprendizaje que, con mayor probabilidad, les lleven a alcanzar los resultados pretendidos (Biggs & Biggs 2005). Una forma de hacerlo es mediante la valoración del conocimiento del estudiante como ser integral, con historia y contexto. A partir de esta perspectiva, el artículo entrega orientaciones para facilitar la adquisición y desarrollo de los procesos de enseñanza y aprendizaje poniendo como protagonista de ellos al estudiante. Pretende ser un aporte en cuanto recurso metodológico de articulación entre las actividades docentes y estudiantiles (de Miguel 2005) como constructores del saber. Lógicamente diferentes formas de abordar el aprendizaje requieren diversidad en la forma de desarrollar la enseñanza, en tanto didáctica como evaluación y su relación con el medio, como elemento fundamental en la significancia del aprendizaje.

Uno de los métodos más extensos y discutidos entre las herramientas estadísticas para el análisis de datos es la regresión. La temática relaciona teóricamente la predicción de una o más variables llamadas *variables dependientes* o *respuesta*, utilizando como base otras variables llamadas *independientes*. En algunos casos las variables independientes reciben otros nombres, *covariables* por ejemplo, también se les denomina *variables explicativas* o *predictores*. Tales problemas son encontrados en casi todas las áreas de la ciencia experimental y la tecnología (Neter et al. 1996, Weisberg 2005).

Cuando el modelo usado para la explicar la variable dependiente en términos de variables independientes asume una relación lineal en los parámetros, se tiene un modelo de regresión lineal; en otro caso se tendrá un modelo de regresión no lineal. Se conoce abundante literatura referente a diferenciar los casos de linealidad y no linealidad, pero ese tópico está fuera del alcance de este artículo (Bates & Watts 1988). Se puntualizará únicamente en los métodos alternativos de regresión.

En la gran mayoría de los textos de regresión lineal simple y múltiple se describen dos métodos para la estimación de parámetros: el método de los mínimos cuadrados y el de máxima verosimilitud (Lehmann & Casella 1998). Estos dos métodos coinciden teóricamente en sus propiedades cuando el modelo que relaciona las variables dependiente e independiente cumple el supuesto de normalidad en los errores, situación que en la realidad difícilmente se presenta y puede llevar a conclusiones erróneas y modelos que no describen apropiadamente la realidad de los datos poblacionales. Estos dos métodos anteriores se conocen como métodos clásicos. No obstante, existen otros métodos menos conocidos, que en algunos textos especializados se encuentran referenciados como temas especiales, algunas veces con poca profundidad teórica y pobre aplicación, generando desinterés por parte del lector.

Los procesos de estimación son variados (Lehmann & Casella 1998) y pueden caracterizar diferentes estimadores y junto a ello un conjunto de criterios que permiten su diferenciación (suficiencia, eficiencia, etc.). En un proceso de aprendizaje, la capacidad discriminativa es fundamental que sea estimulada, de tal manera de permitir al alumno disponer de soportes teóricos que sustenten su opción y superar la justificativa por imposición. Cuando se trabajan los tópicos relativos a regresión lineal, no se debe tener un carácter impositivo del aprendizaje, sino justificado y fundamentado a partir de sus propiedades estadísticas.

El niño en su proceso inicial de aprendizaje conoce o entiende lo que es suave a razón de que conoce lo que es áspero y las posibilidades intermedias. Así, forma parte de su formación la capacidad de discriminación fundamentada (Lopez 2004), pudiendo clasificar futuras situaciones intermedias entre lo suave y lo áspero. De modo similar el concepto de número primo es diferenciable y se justifica su existencia por sus restricciones en divisibilidad que otros números no tienen. La teoría desarrollada en torno al concepto de regresión (mirada convencional) no puede ser diferente y, por tanto, debe existir una modalidad didáctica que complemente el método de enseñanza expositivo o clase magistral, resaltando el esfuerzo metodológico previo y/o paralelo a la formulación formal del concepto de regresión, estimulando la discriminación fundamentada basada en el conocimiento de los métodos no convencionales.

El presente artículo pretende realizar una recopilación de algunos métodos de regresión lineal, no convencionales, basados en la norma L_1 que dan origen a estimadores que son una alternativa en conjuntos de datos que no cumplen los supuestos de normalidad de los errores y describen con mayor claridad la realidad de los mismos.

2. Análisis preliminares

Para poder hablar de regresión, es presentada una modalidad didáctica histórico-evolutiva, que puede complementar el método de enseñanza expositivo o clase magistral, situando en un contexto y resaltando el esfuerzo metodológico previo a la formulación formal del concepto de regresión. La modalidad didáctica que se presenta hace una retrospectiva histórica de lo que antecedió a su descubrimiento.

El término *regresión* fue introducido por Francis Galton en su libro *Natural Inheritance* y fue confirmada por Karl Pearson Madariaga et al. (2013). Su trabajo se centró en la descripción asociativa de los rasgos físicos de los descendientes (variable A) a partir de los de sus padres (variable B). Se observa que el orden fundador de una teoría es un problema concreto, característica estructural de los desarrollos estadísticos.

Estudiando la altura de padres e hijos a partir de más de mil registros de grupos familiares, se llegó a la conclusión de que los padres muy altos tenían una tendencia a tener hijos que heredaban parte de esta altura, pero que revelaban también una tendencia a regresar a la media. Galton generalizó esta tendencia bajo la *ley de la regresión universal*: cada peculiaridad en un hombre es compartida por sus descendientes, pero en media, en un grado menor.

La primera forma de regresión lineal documentada fue el método de los mínimos cuadrados, el cual fue publicado por Legendre en 1805, en *Principle of Least Squares* (Bloomfield & Steiger 1980) donde se incluía una versión del teorema de Gauss-Márkov. Por otro lado, mínimos cuadrados es una técnica de análisis numérico encajada dentro de la optimización matemática, en la que, dados un conjunto de pares (o ternas, etc.), se intenta encontrar la función que mejor se aproxime a los datos, de acuerdo con el criterio de mínimo error cuadrático. En su forma más simple, intenta minimizar la suma de cuadrados de las diferencias ordenadas (llamadas *residuos*) entre los puntos generados por la función y los datos (el modelo observado y el modelo estimado). Específicamente, se llama *mínimos cuadrados promedio* (LMS) cuando el número de datos medidos es 1 y se usa el método de descenso por gradiente para minimizar el residuo cuadrado. Se puede demostrar que LMS minimiza el residuo cuadrado esperado, con el mínimo de operaciones (por iteración), pero requiere un gran número de iteraciones para converger. No obstante, el descubrimiento de los mínimos cuadrados (1755 y 1757, R.J. Boscovitch) articuló un interesante criterio para ajustar una línea a $n \geq 2$ puntos en el plano (Bloomfield & Steiger 1980). Si $(\overline{x}, \overline{y})$ es el centroide de los n

puntos (x_i, y_i), la línea propuesta por Boscovitch escoge c para minimizar:

$$\sum_{i=1}^{n} |y_i - \overline{y} - c(x_i - \overline{x})|$$

Esta es la línea que minimiza el criterio LAD (Least Absolute Deviations) entre todas las líneas restringidas a pasar por la media de los datos. Stigler (1984) propone un algoritmo geométrico para encontrar c, teniendo muchas dificultades computacionales; sin embargo, Laplace ofrece una solución algebraica y elegante, la cual se puede parafrasear de la siguiente manera:

Sin pérdida de generalidad imaginemos $\overline{x} = \overline{y} = 0$ y se observa que la línea LAD a través del origen minimiza:

$$f(c) = \sum_{i=1}^{n} |y_i - cx_i| = \sum_{i=1}^{n} |r_i(c)| \tag{1}$$

Se puede asumir que $x_i \neq 0$ a causa de que $f(c) = \sum_{i=1}^{n} |y_i| + \sum_{i=1}^{n} |y_i - cx_i|$, donde la primera suma es para los $x_i = 0$ y la segunda para los $x_i \neq 0$, entonces f es mínimo cuando la segunda suma lo es.

Ahora imagínese que $\frac{y_i}{x_i} \leq \frac{y_{i+1}}{x_{i+1}}$ pueden ser ordenados en forma ascendente y si c se restringe al intervalo $(\frac{y_p}{x_p}, \frac{y_{p+1}}{x_{p+1}})$, f llega a ser:

$$f(c) = \sum_{i=1}^{p} |x_i| \left(c - \frac{y_i}{x_i}\right) - \sum_{i=p+1}^{n} |x_i| \left(c - \frac{y_i}{x_i}\right) \tag{2}$$

Diferenciar la ecuación anterior permite obtener:

$$f'(c) = \sum_{i=1}^{p} |x_i| - \sum_{i=p+1}^{n} |x_i| \tag{3}$$

Lo cual genera una función lineal a trozos continua con una derivada no decreciente. Si $f' = 0$ para un intervalo $J = (\frac{y_p}{x_p}, \frac{y_{p+1}}{x_{p+1}})$ cualquier c en la clausura de J minimiza (1). Esto permite proponer el siguiente lema:

Lema 1. *f en (1) tiene un minimizador $\hat{c} = \frac{y_i}{x_i}$ para algún $i = 1, \ldots, n$, llámese $i = p$. De esta forma la línea LAD atraviesa el origen conteniendo (x_p, y_p), así al menos un residuo en (1), $r_p(\hat{c})$ es cero.*

El anterior lema motiva un algoritmo:

Algoritmo 1. (1) Calcule $c_i = \frac{y_i}{x_i}$ con $i = 1, \ldots, n$
(2) Evaluar $f(c_i)$ para todo $i = 1, \ldots, n$ y encontrar la función mínima, supóngase $f(c_p)$ y su correspondiente minimizador c_p.

Según Laplace el optimal c es la menor razón en la que la derivada derecha de f es no negativa. Desde (3) se puede ver que $\min\left(j : \sum_{i=1}^{j} |x_i| \geq \sum_{i=j+1}^{n} |x_i|\right)$. En otras palabras:

$$p = \min\left(j : \sum_{i=1}^{j} |x_i| \geq \sum_{i=j+1}^{n} |x_i|/2\right) \tag{4}$$

El valor $c = \frac{y_p}{x_p}$ es la mediana ponderada de los $\frac{y_i}{x_i}$ con pesos $|x_i|$, y la mediana ponderada puede ser obtenida en aproximadamente un tiempo proporcional a $nlog(n)$.

Lema 2. *La línea LAD a través del origen es la mediana ponderada de los* $\frac{y_i}{x_i}$ *con pesos* $|x_i|$ (x_i, y_i)*; son aquellos puntos para los cuales* $x_i \neq 0$*. la complejidad esperada no es mayor de* $O(nlog(n))$*.*

3. Modalidad didáctica

3.1. Métodos de regresión alternativa

Los siguientes métodos constituyen una alternativa para los modelos de regresión convencional.

3.1.1. Regresión MINMAD

Charnes et al. (1955) proponen la necesidad de utilizar la minimización de la norma L_1, para ajustar un modelo simple, en una situación donde no es posible utilizar mínimos cuadrados, con el fin de determinar el porcentaje de distintos factores aplicados, para determinar el salario de los ejecutivos en una empresa del sector industrial. Es decir, dos siglos después de que R.J. Boscovitch propusiera la recta de ajuste LAD (Least Absolute Deviations), estos autores deciden retomar esta metodología. En este artículo los autores proponen que se realice la estimación utilizando el *Método Simplex*. Tres años más tarde Karst (1958) propone una metodología estadística para encontrar la solución a este problema.

Se trata de minimizar la media de los valores absolutos de las desviaciones de las observaciones con respecto a la línea de regresión, de ahí su nombre MINMAD (Minimizing Mean of Absolute Deviations), en un modelo de regresión lineal simple, a diferencia de mínimos cuadrados que minimiza la norma cuadrática o euclidiana. Es decir, ahora se trata de estimar β_0 y β_1 minimizando:

$$\frac{1}{n}\sum_{i=1}^{n} |Y_i - \beta_0 - \beta_1 X_i| \tag{5}$$

Esto equivale a minimizar:

$$\sum_{i=1}^{n} |Y_i - \beta_0 - \beta_1 X_i| \tag{6}$$

Para simplificar el análisis, primeramente se imponen restricciones a β_0 y β_1 para que satisfagan la condición $Y_0 = \beta_0 + \beta_1 X_0$ para un par dado (X_0, Y_0). Además, dado el par (X_0, Y_0), es posible transformar los datos:

$$\begin{aligned} x_i =& X_i - X_0 \\ y_i =& Y_i - Y_0 \end{aligned}$$

Luego al reemplazar los datos transformados en (28), se tiene:

$$\sum_{i=1}^{n} |Y_i - \beta_0 - \beta_1 X_i| = \sum_{i=1}^{n} |y_i + Y_0 - \beta_0 - \beta_1 (x_i + X_0)|$$

$$\sum_{i=1}^{n} |y_i + \beta_0 + \beta_1 X_0 - \beta_0 - \beta_1 x_i - \beta_1 X_0| = \sum_{i=1}^{n} |y_i - \beta_1 x_i|$$

Así, el problema se reduce ahora a calcular un β que minimice la expresión

$$\sum_{i=1}^{n} |y_i - \beta x_i| \tag{7}$$

Por ejemplo, se usarán los siguientes tres datos:

i	x_i	y_i
1	1	3
2	1	1
3	2	4

Ver figura 1.

Teorema 1. *La función $f(\beta) = \sum_{i=1}^{n} |y_i - \beta x_i|$, para los valores dados de (x_i, y_i), con $i = 1, \cdots, n$ es una función lineal a trozos y convexa.*

Demostración:
Se debe demostrar que para $\beta' < \beta''$ y $0 \leq \lambda \leq 1$, y $\beta = \lambda\beta' + (1 - \lambda)\beta''$:

$$f(\beta) \leq \lambda f(\beta') + (1 - \lambda) f(\beta'')$$

En efecto, se sabe que $f_i(\beta) = |y_i - \beta x_i|$

$$f_i(\beta) = f_i(\lambda\beta' + (1 - \lambda)\beta'') = |y_i - \lambda\beta' x_i - (1 - \lambda)\beta'' x_i|$$

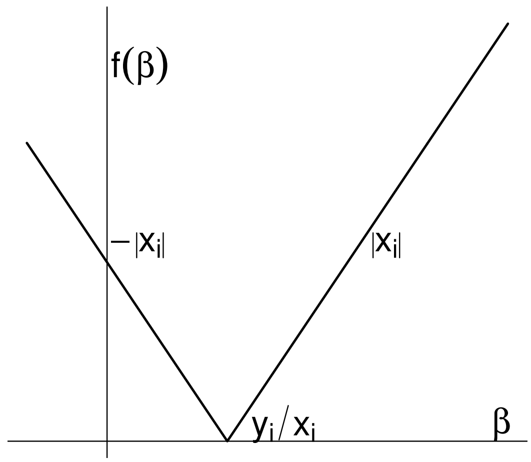

Figura 1: *Visualización de* $|y_i - \beta x_i|$ *para cualquier* i *caracterizando dos líneas rectas con un mínimo en* $(\frac{y_i}{x_i}, 0)$ *y pendientes* $-|x_i|$ *y* $|x_i|$. *Fuente: elaboración propia.*

$$\begin{aligned}
&= |\lambda(y_i - \beta' x_i) + (1-\lambda)(y_i - \beta'' x_i)| \\
&\leq \lambda |y_i - \beta' x_i| + (1-\lambda)|y_i - \beta'' x_i| \\
&= \lambda f_i(\beta') + (1-\lambda) f_i(\beta'')
\end{aligned}$$

Sabiendo que la suma de dos funciones convexas es convexa, se tiene que $f(\beta)$ es convexa. Además, sabiendo que $f_i(\beta)$ es lineal a trozos, la suma de funciones lineales a trozos es lineal a trozos; en consecuencia, se obtiene el resultado (ver figura 2).

Teorema 2. *La función* $f(\beta)$ *tiene las siguientes propiedades:*

1. *La pendiente del segmento del extremo de la izquierda es* $-\sum_{i=1}^{n} |x_i|$ *, y de la derecha es* $\sum_{i=1}^{n} |x_i|$.

2. *Los vértices de la función poligonal* $f(\beta)$ *son de la forma* $\left(\frac{y_i}{x_i}\right)$ *, donde* $\frac{y_i}{x_i}$ *es el mínimo de* $f_i(\beta)$. *Si* $(i_1, \cdots, i_n)$ *es un conjunto de índices tales que* $\frac{y_{i_1}}{x_{i_1}} \leq \ldots \leq \frac{y_{i_n}}{x_{i_n}}$, *entonces la pendiente de* $f(\beta)$ *se incrementa en* $2|x_{i_k}|$ *al pasar por* $\beta_k = \frac{y_{i_k}}{x_{i_k}}$.

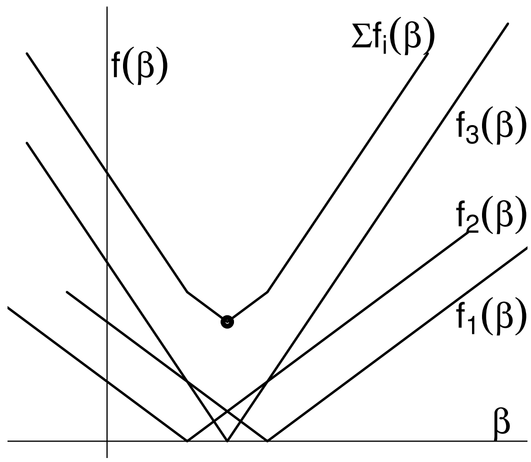

Figura 2: $|y_i - \beta x_i|$ *para* $i = 1, 2, 3$ *y* $\sum |y_i - \beta x_i|$ *es una función lineal a trozos y convexa. Fuente: elaboración propia.*

Estos resultados proporcionan un método para calcular el mínimo de $f(\beta)$. El mínimo se alcanza en un β_r tal que:

$$\begin{aligned} -\sum_{i=1}^{n} |x_i| + 2\sum_{k=1}^{r-1} |x_{i_k}| &< 0 \\ -\sum_{i=1}^{n} |x_i| + 2\sum_{k=1}^{r} |x_{i_k}| &\geq 0 \end{aligned} \tag{8}$$

Si $-\sum_{i=1}^{n} |x_i| + 2\sum_{k=1}^{r} |x_{i_k}| = 0$ entonces $\beta_{(r)} \leq \beta \leq \beta_{(r+1)}$ son optimales. Entonces se puede escoger $\beta_{(r)}$ o $\beta_{(r+1)}$ con igual probabilidad. Los parámetros estimados en este caso son:

$$\begin{aligned} \widetilde{\beta}_1 &= \frac{y_r}{x_r} \\ \widetilde{\beta}_0 &= Y_0 - \left(\frac{y_r}{x_r}\right) X_0 \end{aligned} \tag{9}$$

Si se escoge $\beta_{(r)}$ como solución. La teoría anterior permite generar un algoritmo:

Algoritmo 2. Dados los puntos muestrales (X_i, Y_i) para $i = 1, \ldots, n$ de (X, Y) calcular:

1. $(X_0, Y_0) = (\overline{X}, \overline{Y})$.

2. Calcular las variables transformadas $x_i = X_i - X_0$ y $y_i = Y_i - Y_0$.

3. Calcular los mínimos de las funciones f_i como $\frac{y_i}{x_i}$.

4. Asignar rangos a los mínimos anteriores en forma ascendente, dando el valor 1 al más pequeño y n al más grande.

5. Calcular $-\sum_{i=1}^{n} |x_i|$ y comience a sumar $2|x_i|$, siguiendo el orden de los rangos.

6. Cuando la suma anterior pase de negativa a positiva, escójase el mínimo $\frac{y_i}{x_i}$ en ese paso como estimador de β_1. Si esto ocurre en el paso r, entonces $\widetilde{\beta}_1 = \frac{y_r}{x_r}$ y calcúlese $\widetilde{\beta}_0$.

Birkes & Dodge (2011) proponen un algoritmo, que aunque parezca eficiente, es demasiado lento y necesita muchas tablas para ser comparadas, lo cual a su vez implica muchas iteraciones.

Algoritmo 3. El objetivo de este algoritmo es encontrar la mejor línea de ajuste entre todas las líneas. Dado un punto (X_0, Y_0) de los datos, para cada punto (X_i, Y_i) calcular:

1. La pendiente $\frac{(Y_i - Y_0)}{(X_i - X_0)}$ de la línea pasando a través de los dos puntos (X_0, Y_0) y (X_i, Y_i). Si $X_i = X_0$ para algún i, tales puntos pueden ser ignorados.

2. Reindexar los puntos tal que: $\frac{(Y_1 - Y_0)}{(X_1 - X_0)} \leq \frac{(Y_2 - Y_0)}{(X_2 - X_0)} \leq \cdots \leq \frac{(Y_n - Y_0)}{(X_n - X_0)}$. Sea $T = |X_i - X_0|$.

3. Encontrar el índice k que satisfaga las condiciones:

$$|X_i - X_0| + \cdots |X_{k-1} - X_0| < \frac{1}{2}T$$
$$|X_i - X_0| + \cdots |X_{k-1} - X_0| + |X_k - X_0| > \frac{1}{2}T$$

4. La mejor línea pasando a través de (X_0, Y_0) es la línea $\hat{Y} = {\beta_0}^* + {\beta_1}^* X$, donde:

$${\beta_1}^* = \frac{Y_k - Y_0}{X_k - X_0}$$
$${\beta_0}^* = Y_0 - {\beta_1}^* X_0$$

Para más detalles se recomienda leer Sanjith & Elangovan (2014).

3.1.2. Regresión MINMAXAD

Este método consiste en estimar los parámetros β_0 y β_1 minimizando la máxima de las desviaciones absolutas (Minimizing Maximum of Absolute Deviations). Bajo este criterio, la función objetivo es:

$$\mathrm{Min}_{(\beta_0,\beta_1)}\left[\mathrm{Máx}\left|Y_i-\beta_0-\beta_1X_i\right|\right]$$

De momento se discute el problema sin el término β_0. Si $f_i(\beta)=|Y_i-\beta X_i|$ y $g(\beta)=\mathrm{máx}_i f_i(\beta)$, se puede demostrar que $g(\beta)$ es una función lineal a trozos convexa. Los vértices de $g(\beta)$ no son necesariamente los puntos $\frac{Y_i}{X_i}$ como ocurría antes. Los vértices de $g(\beta)$ son las coordenadas β de las intersecciones de las líneas $g(\beta)=Y_i+X_i\beta$ o $g(\beta)=-(Y_i+X_i\beta)$ con las líneas $g(\beta)=Y_j+X_j\beta$ o $g(\beta)=-(Y_j+X_j\beta)$ para cada $i\neq j$.

La interpretación geométrica de este método es ampliamente discutida por Wagner (1959), quien propone un algoritmo para variables acotadas, y por Stiefel (1960), quien reduce el problema al método Simplex. La figura 3 ilustra la idea geométrica, se pretende encontrar el mínimo de la función $g(\beta)$. Para más detalles se recomienda leer a Popescu & Supian (2007)

3.1.3. MINSADBED

Este tipo de regresión considera la estimación de los parámetros β_0 y β_1 minimizando la suma de las diferencias absolutas entre desviaciones (Minimizing Sum of Absolute Difference Betwen Deviations), esto es:

$$\text{Minimizar}\sum_{i<j}|d_i-d_j|$$

Es de aclarar que las distancias d_i y d_j representan las diferencias entre el modelo observado Y_i y el modelo a ser estimado $\beta_0+\beta_1X_i$ en el primer caso y $\beta_0+\beta_1X_j$ en el segundo. Así:

$$\begin{aligned}\sum_{i<j}|d_i-d_j| &= \sum_{i<j}|(Y_i-\beta_0-\beta_1X_i)-(Y_j-\beta_0-\beta_1X_j)|\\ &= \sum_{i<j}|(Y_i-Y_j)-\beta_1(X_i-X-j)\end{aligned}$$

Haciendo $Y_{ij}=Y_i-Y_j$ y $X_{ij}=X_i-X_j$, se tiene:

$$\sum_{i<j}|d_i-d_j|=\sum_{i<j}|Y_{ij}-\beta_1X_{ij}|$$

Este problema se reduce al problema MINMAD, excepto que se deben realizar las diferencias entre Y_i y Y_j y entre X_i y X_j, las cuales suman un total de $\frac{n(n-1)}{2}$. El parámetro β_0 se calcula haciendo:

$$\hat{\beta}_0=\overline{Y}-\overline{X}\hat{\beta}_1$$

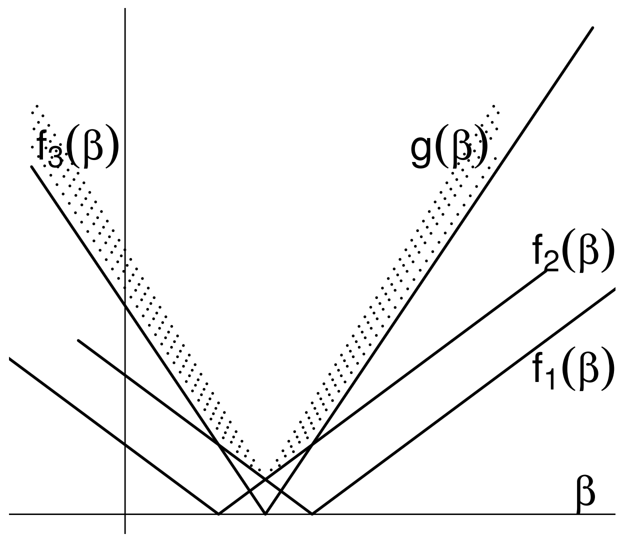

Figura 3: *Idea geométrica de la regresión MINMAXAD. Fuente: elaboración propia.*

Y otra forma alternativa de estimarlo es:

$$\hat{\beta}_0 = \text{Mediana}_{i<j} \frac{1}{2}(Y_i + Y_j)$$

Para más detalles ver Sánchez (2014).

3.1.4. Regresión MINSADBAD

Este método consiste en estimar los parámetros β_0 y β_1 minimizando la suma de las diferencias absolutas entre desviaciones absolutas (Minimizing Sum of Absolute Differences Between Absolute Deviations). Es decir:

$$\text{Minimizar}_{\beta_0,\beta_1} \sum_{i<j} ||d_i| - |d_j||$$

Donde $d_i = Y_i - (\beta_0 + \beta_1 X_i)$. Este método resulta ser más apropiado explicarlo, cuando se considera el modelo de regresión lineal múltiple. Para más detalles se recomienda leer Sanjith & Elangovan (2014).

4. Conclusiones

La forma de abordar los procesos de enseñanza y aprendizaje ha transitado desde un modelo centrado en la enseñanza a uno centrado en el aprendizaje del estudiante y su relación con factores personales y contextuales. Este cambio de paradigma invita a cuestionarse qué se pretende que aprendan los estudiantes, cuáles son las modalidades y metodologías más adecuadas para que puedan desarrollar y adquirir estos aprendizajes y con qué criterios y procedimientos vamos a comprobar si el estudiante los ha adquirido finalmente y la transferencia que pueda hacer de ellos. Pero también es una invitación a apoyar a los profesores, quienes en su mayoría fueron formados en el paradigma donde lo más importante era el contenido y la enseñanza de estos, por lo que requieren modelos y estrategias de enseñanza que faciliten la implementación del modelo de enseñanza centrado en el aprendizaje del estudiante.

La teoría desarrollada en torno al concepto de regresión (mirada convencional) no puede ser diferente y, por tanto, debe existir una modalidad didáctica que complemente el método de enseñanza expositivo o clase magistral, resaltando el esfuerzo metodológico previo y/o paralelo a la formulación formal del concepto de regresión, estimulando la discriminación fundamentada basada en el conocimiento de los métodos no convencionales. Se considera prioritario impulsar estudios que faciliten estrategias a los profesores universitarios para promover en las aulas un nuevo enfoque de los procesos de enseñanza y aprendizaje. En este artículo se intenta facilitar algunas orientaciones que permitan la concreción de los procesos de enseñanza y aprendizaje de la regresión, teniendo en cuenta las implicaciones metodológicas que tiene este cambio de paradigma. El presente artículo presentó una recopilación de algunos métodos de regresión lineal, no convencionales, basados en la norma L_1, que dan origen a estimadores que son una alternativa en conjuntos de datos que no cumplen los supuestos de normalidad de los errores y describen con mayor claridad la realidad de estos como una modalidad didáctica histórico-evolutiva de enseñanza de la regresión.

Recibido: 12 de enero de 2016
Aceptado: 3 de marzo de 2016

Referencias

Bates, D. M. & Watts, D. G. (1988), *Nonlinear regression: iterative estimation and linear approximations*, Wiley Online Library.

Biggs, J. & Biggs, J. B. (2005), *Calidad del aprendizaje universitario*, Vol. 7, Narcea Ediciones.

Birkes, D. & Dodge, Y. (2011), *Alternative methods of regression*, Vol. 190, John Wiley & Sons.

Bloomfield, P. & Steiger, W. (1980), 'Least absolute deviations curve-fitting', *SIAM Journal on Scientific and Statistical Computing* **1**(2), 290–301.

Buendía, L. & Olmedo, E. M. (2002), 'El género:¿ constructo mediador en los enfoques de aprendizaje universitario?', *Revista de Investigación Educativa* **20**(2), 511–524.

Charnes, A., Cooper, W. W. & Ferguson, R. O. (1955), 'Optimal estimation of executive compensation by linear programming', *Management Science* **1**(2), 138–151.

de Miguel, M. (2005), 'Modalidades de enseñanza centradas en el desarrollo de competencias', *Orientaciones para promover el cambio metodológico en el Espacio Europeo de Educación Superior. Oviedo: Universidad de Oviedo* pp. 109–14.

Entwistle, N. & Ramsden, P. (2015), *Understanding Student Learning (Routledge Revivals)*, Routledge.

Karst, O. J. (1958), 'Linear curve fitting using least deviations', *Journal of the American Statistical Association* **53**(281), 118–132.

Lehmann, E. L. & Casella, G. (1998), *Theory of point estimation*, Vol. 31, Springer Science & Business Media.

Lopez (2004), 'La toma de decisiones en los sistemas de autoayuda y asesoramiento vocacional', *Universidad Complutense de Madrid* **50**(3), 1–9.

Madariaga, D. F. C., Rodríguez, J. L. G., Lozano, M. R., Vallejo, E. C. & de Administración, E. (2013), 'Inferencia estadística módulo de regresión lineal simple'.

Neter, J., Kutner, M. H., Nachtsheim, C. J. & Wasserman, W. (1996), *Applied linear statistical models*, Vol. 4, Irwin Chicago.

Popescu, C. C. & Supian, S. (2007), 'An application of minsadbesd regression', *Journal of Applied Quantitative Methods* p. 160.

Ramsden, P. (2003), *Learning to teach in higher education*, Routledge.

Sánchez, A. (2014), 'Comparación de algunos métodos de regresión alternativa vs. estadística bayesiana usando mcmc'.

Sanjith & Elangovan (2014), 'Leas absolute deviations estimation for the censored regression model', *Asia Pacific Journal of Research* **1**, XVI.

Stiefel, E. (1960), 'Note on jordan elimination, linear programming and tchebycheff approximation', *Numerische Mathematik* **2**(1), 1–17.

Stigler, S. M. (1984), 'Studies in the history of probability and statistics xl boscovich, simpson and a 1760 manuscript note on fitting a linear relation', *Biometrika* **71**(3), 615–620.

Wagner, H. M. (1959), 'Linear programming techniques for regression analysis', *Journal of the American Statistical Association* **54**(285), 206–212.

Weisberg, S. (2005), *Applied linear regression*, Vol. 528, John Wiley & Sons.