UNIVERSIDAD SANTO TOMAS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

Comunicaciones en Estadística
August 2016, Vol. 9, No. 2, pp. 211–224

Historical-Evolutionary teaching mode for regression [1]

Modalidad didáctica histórico-evolutiva para la regresión

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

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

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

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 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.

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

[1]Sánchez, A., Sarzoza, S., González J. Historical-Evolutionary teaching mode for regression *Comunicaciones en Estadística*, **9**(2), 211-224.

[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.

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.

1 Introduction

It is not enough to say that teachers must change their role when planning their teaching, but it is necessary to introduce models and guidelines that help them in this process. Precisely because we want to avoid this risk, it is necessary to promote among university professors studies and models oriented to generate a favorable culture towards the paradigm change in the teaching processes, and it should be accompanied by instrumental procedures that make easier the approaching to this task. And this is exactly what is proposed in this paper, that for a set of technical orientations is established, which allows the professors to distribute their teaching activity in different modalities and teaching and learning methodologies (de Miguel 2005).

It is imperative to promote studies that facilitate strategies for university professors, in order to encourage the acquisition of learning from the interaction of factors, both personal and contextual. It is all about the direct relation between the teaching method, the evaluation model and the way in which the student approaches learning from his own perspective (Entwistle & Ramsden 2015, Ramsden 2003, Biggs & Biggs 2005, Buendía & Olmedo 2002).

The way in which the student faces the learning situations is influenced by his perception about the means and its relation to teaching, which is defined by the conceptions of the professors about it and the implications for the learning of their students. For students to achieve the desired results and in a reasonably efficient manner, the fundamental task of the teacher consists in making his students to perform the learning activities that, most likely, will lead to achieve the intended results (Biggs & Biggs 2005). A way of doing it is through the appreciation of the student's knowledge as an integral being, with history and context. From this perspective, this paper gives orientations to facilitate the acquisition and development of the teaching and learning processes turning the student into a protagonist of them. It pretends to be a contribution as a methodological resource for the articulation between the teaching activities and the student activities (de Miguel 2005) as builders of knowledge. Naturally, different approaches require diversity in the way of developing teaching, as didacticism as assessment and its relationship with the environment as a major factor in the meaning of learning.

One of the most comprehensive and discussed methods among statistical tools for the data analysis is the regression model. The thematic relates theoretically the prediction of one or more variables called dependent variables or response, using as base other variables called independent. In some cases the independent variables receive other names, covariables, for example, they are also called explanatory variables or predictors. Such problems are found in almost all areas of experimental science and technology (Neter et al. 1996, Weisberg 2005).

When the model used to explain the dependent variable in terms of independent variables assumes a linear relation in the parameters, we have a linear regression model; in another case, we would have a non-linear regression model. There is abundant literature about the difference between the cases of linearity and no linearity, but this topic is out of reach in this paper (Bates & Watts 1988). We will point out only alternative regression methods.

In the vast majority of the texts of simple linear and multiple regression, two methods for estimating parameters are described: the least squares and maximum verisimilitude method (Lehmann & Casella 1998). These two methods coincide theoretically in their properties when the model relating the dependent and independent variables meet the assumption of normality in the errors, a situation that rarely occurs in reality and can lead to erroneous conclusions and models that do not adequately describe the reality of population data. These two previous methods are known as classical methods. However, there are other lesser known methods, which in some specialized texts are referenced as special topics, sometimes with little theoretical depth and poor application, generating disinterest to the reader.

Estimation processes are varied (Lehmann & Casella 1998) and may characterize different estimators and along with it a set of criteria that allow their differentiation (sufficiency, efficiency, etc.). To stimulate discriminative capacity is essential in a learning process because it allows the students having theoretical supports to back their options and get over the justificatory by imposition. To handle topics about linear regression it's necessary to get rid of the imposing character of learning and adopt a justified character based on their statistical properties.

Children in their initial process of learning know and understand what is soft because they know what is hard and the intermediate possibilities. So, substantiated discrimination is part of their education (Lopez 2004), being able to classify future intermediate situations between the soft and the hard. In a similar way, the concept of prime number is distinguishable and its existence is justified by its restrictions in divisibility that other numbers do not have. The developed theory around the idea of regression (conventional view) can not be different and, therefore, it must be a teaching modality that complements the expository method or master class, highlighting the methodological effort and/or parallel to the formal formulation of the concept of regression, stimulating the discrimination based on the knowledge of the non-conventional methods.

This paper aims to make a compilation of some non-conventional methods of linear regression based on the L_1 norm that gives rise to estimators, which are an alternative to sets of data that don't meet the assumptions of normality of errors and describe with more clarity the reality of the same

2 Preliminary analysis

To talk about regression, we introduce a teaching historical-evolutionary modality that may complement the expository or masterly teaching method, locating in a context and highlighting the methodological effort previous to the formal formulation of the concept regression. The teaching modality introduced makes a historical retrospective of what preceded its discovering.

The term *regression* was introduced by Francis Galton in his book *Natural Inheritance* and was confirmed by Karl Pearson Madariaga et al. (2013). His work was focused on the associative description of the physical features of the descendants (variable A) from their fathers' (variable B). It is noted that the founding order of a theory is a particular problem, structural feature of the statistical developments.

Studying the height of parents and their children from more than one thousand registers of family groups, it was concluded that the very tall parents had a tendency to have children that inherited part of that height, but it also revealed a tendency to return to the mean. Galton generalized this trend under the *law of universal regression*; each particularity in a man is shared by his descendants, but in average, in a minor grade.

The first way of linear regression documented was the method of least squares, which was published by Legendre in 1805, in *Principle of Least Squares* (Bloomfield & Steiger 1980) where a version of the Gauss-Markov theorem was included. On the other hand, least squares is a technique of numerical analysis embedded within the mathematical optimization, in which, given a set of pairs (or shortlists) it is intended to find the function that better approaches to data, according to the criteria of least quadratic error. In its simplest way, it tries to minimize the sum of squares of the orderes differences (called *residuals*) among the points generated by the function and the data (the observed model and the fitted model). Specifically, it is called *least mean squares* (LMS) when the number of data measured is 1, and the gradient descent method is used to minimize the residual squares. It can be proved that LMS minimizes the expected residual squares , with the minimum of operations (by iteration), but it requires a big number of iterations to converge. However, the discovering of the least squares (1755 y 1757, R.J. Boscovitch) articulated an interesting criterion to fit a line to $n \geq 2$ points in the plane (Bloomfield & Steiger 1980). If $(\overline{x}, \overline{y})$ is the centroid of the n points (x_i, y_i), the line proposed by Boscovitch chooses c to minimize:

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

This is the line that minimizes the LAD criteria (Least Absolute Deviations) among all the lines restricted to pass through the mean of the data. Stigler (1984) proposes a geometrical algorithm to find c, having many computational difficulties; however, Laplace offers an algebraic and elegant solution, which can be paraphrased as following:

Without loss of generality suppose that $\overline{x} = \overline{y} = 0$ and we observe that the LAD line passing through the origin minimizes:

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

It can be assumed that $x_i \neq 0$ because $f(c) = \sum_{i=1}^{n} |y_i| + \sum_{i=1}^{n} |y_i - cx_i|$, where the first sum is for all $x_i = 0$ and the second for the $x_i \neq 0$, then f is minimal when the second sum is.

Now, imagine that $\frac{y_i}{x_i} \leq \frac{y_{i+1}}{x_{i+1}}$ can be arranged in an ascendant way and if c is restricted to the interval $(\frac{y_p}{x_p}, \frac{y_{p+1}}{x_{p+1}})$, f becomes:

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

Differentiate the previous equation allows obtaining:

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

Which generates a piecewise continuous linear function with a non-decreasing derivate. If $f' = 0$ for an interval $J = (\frac{y_p}{x_p}, \frac{y_{p+1}}{x_{p+1}})$, any c in the closure of J minimizes (1). It allows proposing the following lema:

lema 1. *f defined in (1) has a minimizer $\hat{c} = \frac{y_i}{x_i}$ for some $i = 1, \ldots, n$, be it called $i = p$. In this way the LAD line passes through the origin containing (x_p, y_p), so at least one residue in (1) $r_p(\hat{c})$ is zero.*

Previous motto motivates the following algorithm:

Algorithm 1. (1) Calculate $c_i = \frac{y_i}{x_i}$ with $i = 1, \ldots, n$
(2) Evaluate $f(c_i)$ for all $i = 1, \ldots, n$ and find the minimal value, named $f(c_p)$ and its correspondent minimizer, c_p.

According to Laplace, the optimal c is the minor rate in which the right derivate of f is non-negative. From (3) we can see that $\min\left(j : \sum_{i=1}^{j} |x_i| \geq \sum_{i=j+1}^{n} |x_i|\right)$. In other words:

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

The value $c = \frac{y_p}{x_p}$ is the weighted median of the $\frac{y_i}{x_i}$ with weights $|x_i|$, and the weighted median can be obtained in approximately a time proportional to $nlog(n)$.

lema 2. *The LAD line passing through the origin is the weighted median of $\frac{y_i}{x_i}$ with weights $|x_i|$ (x_i, y_i); are those points for which $x_i \neq 0$. The expected complexity is not bigger than $O(nlog(n))$.*

3 Teaching modality

3.1 Methods of alternative regression

Following methods are an alternative for the model of conventional regression.

3.1.1 MINMAD Regression

Charnes et al. (1955) propose the need of using the minimization of the L_1 norm to fit a simple model, in a situation where it is impossible to use least squares, to determinate the percentage of different applied factors, to determine the salary of executives in a company of the industry field. Two centuries after R.J Boscovitch proposed the fit of LAD line (Least Absolute Deviations) these authors decide to resume this methodology. In this paper, authors propose to carry out the estimation using the *Simplex method*. Three years later, Karst (1958) proposes a statistical methodology to find the solution to this problem.

It is about minimizing the mean of the absolute values of the deviations of observations regarding the line of regression, that's why it is named MINMAD (Minimizing Mean of Absolute Deviations), in a model of simple linear regression, different to least squares that minimizes the quadratic or Euclidian norm. It is about estimating β_0 and β_1 minimizing:

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

This is equivalent to minimize:

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

To simplify the analysis, firstly, restrictions are imposed to β_0 and β_0 so they satisfy the condition $Y_0 = \beta_0 + \beta_1 X_0$ for a pair given (X_0, Y_0). Besides given the pair (X_0, Y_0) it's possible to transform the data:

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

Then, when replacing the transformed data in (28), we have:

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

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

So, the problem is reduced now to calculate a β that minimizes the expression

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

For example the following three data will be used:

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

See Figure 1.

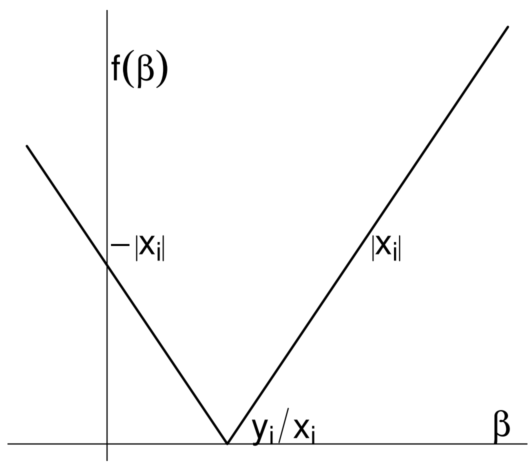

Figure 1: *Visualization of $|y_i - \beta x_i|$ for any i characterizing two straight lines with minimum at $(\frac{y_i}{x_i}, 0)$ and slopes $-|x_i|$ y $|x_i|$. Source: own elaboration.*

Theorem 1. *The function $f(\beta) = \sum_{i=1}^{n} |y_i - \beta x_i|$ for the given values of (x_i, y_i), with $i = 1, \cdots, n$ is a piecewise linear convex function.*

Proof:
It must be proved that for $\beta' < \beta''$, $0 \leq \lambda \leq 1$, and $\beta = \lambda\beta' + (1 - \lambda)\beta''$:

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

Indeed, it is known that $f_i(\beta) = |y_i - \beta x_i|$

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

Knowing that the sum of two convex functions is convex, we have that $f(\beta)$ is convex. Besides, knowing that $f_i(\beta)$ is piecewise linear, the sum of linear piecewise functions is piecewise linear; as a consequence, we obtain the result (see figure 2).

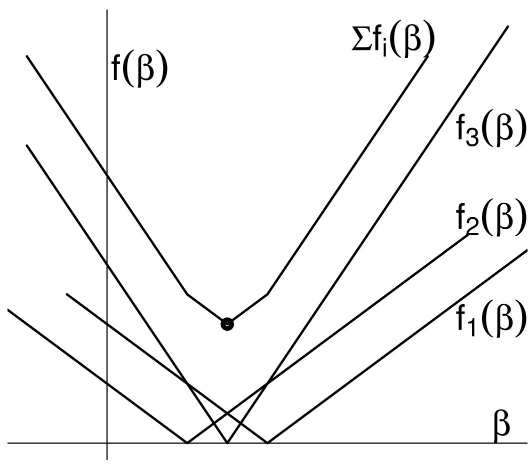

Figure 2: $|y_i - \beta x_i|$ *for* $i = 1, 2, 3$ *and* $\sum |y_i - \beta x_i|$ *is a linear piecewise convex function. Source: own elaboration.*

Theorem 2. *The function* $f(\beta)$ *has the following properties:*

The slope of the segment of the leftmost is $-\sum_{i=1}^{n} |x_i|$ *, and the one of the right is* $\sum_{i=1}^{n} |x_i|$.

The vertices of the polygonal function $f(\beta)$ *are of the shape* $\left(\frac{y_i}{x_i}\right)$ *, where* $\frac{y_i}{x_i}$ *is the minimum of* $f_i(\beta)$*. If* $(i_1, \cdots, i_n)$ *is a set of indexes such that* $\frac{y_{i_1}}{x_{i_1}} \leq \ldots \leq \frac{y_{i_n}}{x_{i_n}}$*, then the slope of* $f(\beta)$ *increases in* $2|x_{i_k}|$ *when passing through* $\beta_k = \frac{y_{i_k}}{x_{i_k}}$.

These results provide a method to calculate the minimum of $f(\beta)$. The minimum is reached in a β_r such that:

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

If $-\sum_{i=1}^{n}|x_i| + 2\sum_{k=1}^{r}|x_{i_k}| = 0$ then $\beta_{(r)} \leq \beta \leq \beta_{(r+1)}$ are optimal. Then we can choose $\beta_{(r)}$ or $\beta_{(r+1)}$ with the same probability. Estimated parameters in this case are:

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

If we choose $\beta_{(r)}$ as solution. The previous theory allows generating an algorithm:

Algorithm 2. Given the sampling points (X_i, Y_i) with $i = 1, \ldots, n$ of (X, Y), calculate:

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

Calculate transformed variables $x_i = X_i - X_0$ and $y_i = Y_i - Y_0$.

Calculate the minimums of the functions f_i as $\frac{y_i}{x_i}$.

Assign ranks to the previous minimums in an ascendant way, giving the value 1 to the smallest and n to the biggest

Calculate $-\sum_{i=1}^{n}|x_i|$ and start summing $2|x_i|$, following the ranks' order.

When previous sum turns from negative to positive, choose the minimum $\frac{y_i}{x_i}$ in that step as estimator of β_1. If this taks place in the step r, then $\widetilde{\beta}_1 = \frac{y_r}{x_r}$ and calculate $\widetilde{\beta}_0$.

Birkes & Dodge (2011) propose an algorithm that, although it seems efficient, it is too slow and needs many tables to be compared, which in turn implies many iterations.

Algorithm 3. The aim of this algorithm is to find the best-fitted line among all lines. Given a point (X_0, Y_0) of the data, for each point (X_i, Y_i) calculate:

The slope of the line $\frac{(Y_i - Y_0)}{(X_i - X_0)}$ passing through both points (X_0, Y_0) and (X_i, Y_i). If $X_i = X_0$ for some i, such points can be ignored.

Reindex the points such that: $\frac{(Y_1 - Y_0)}{(X_1 - X_0)} \leq \frac{(Y_2 - Y_0)}{(X_2 - X_0)} \leq \cdots \leq \frac{(Y_n - Y_0)}{(X_n - X_0)}$. Define $T = |X_i - X_0|$.

Find the index k that satisfies the conditions:

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

The best line passing through (X_0, Y_0) is the line $\hat{Y} = {\beta_0}^* + {\beta_1}^* X$, where:

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

For more details, we recommend reading Sanjith & Elangovan (2014).

3.1.2 MINMAXAD Regression

This method consists in estimating the parameters β_0 and β_1 minimizing the maximum absolute deviations. Under this criterion, the target function is:

$$\text{Min}_{(\beta_0,\beta_1)}\Big[\text{Máx}\big|Y_i - \beta_0 - \beta_1 X_i\big|\Big]$$

At the moment the problem is discussed without the term β_0. If $f_i(\beta) = |Y_i - \beta X_i|$ and $g(\beta) = \text{máx}_i f_i(\beta)$, it can be proved that $g(\beta)$ is a convex piecewise linear function. The vertices of $g(\beta)$ are not necessarily the points $\frac{Y_i}{X_i}$ as would happen before. The vertices of $g(\beta)$ are the coordinates β of the intersections of the lines $g(\beta) = Y_i + X_i\beta$ or $g(\beta) = -(Y_i + X_i\beta)$ with the lines $g(\beta) = Y_j + X_j\beta$ or $g(\beta) = -(Y_j + X_j\beta)$ for each $i \neq j$.

The geometrical interpretation of this method is extensively discussed by Wagner (1959), who proposes an algorithm for bounded variables, and by Stiefel (1960), who reduces the problem to the Simplex method. Figure 3 shows the geometrical idea; it is intended to find the minimum of the function $g(\beta)$. For more details, it's recommended to read Popescu & Supian (2007).

3.1.3 MINSADBED

This kind of regression considers the estimation of the parameters β_0 and β_1 minimizing the sum of the absolute differences between deviations, it is:

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

It is to clarify that the distances d_i and d_j represent the differences between the observed model Y_i and the model to be estimated $\beta_0 + \beta_1 X_i$ in the first case, and $\beta_0 + \beta_1 X_j$ in the second. This way:

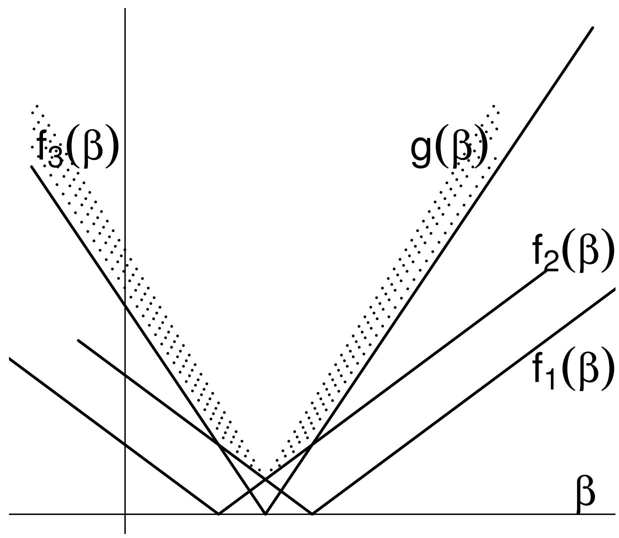

Figure 3: *Geometrical idea of the MINMAXAD regression. Source: own elaboration.*

$$\sum_{i<j} |d_i - d_j| = \sum_{i<j} |(Y_i - \beta_0 - \beta_1 X_i) - (Y_j - \beta_0 - \beta_1 X_j)|$$
$$= \sum_{i<j} |(Y_i - Y_j) - \beta_1 (X_i - X - j)$$

Defining $Y_{ij} = Y_i - Y_j$ and $X_{ij} = X_i - X_j$, we have:

$$\sum_{i<j} |d_i - d_j| = \sum_{i<j} |Y_{ij} - \beta_1 X_{ij}|$$

This problem is reduced to the MINMAD problem, except that the differences between Y_i and Y_j and between X_i and X_j , which sum a total of $\frac{n(n-1)}{2}$. Parameter β_0 is calculated by:

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

Another alternative way to estimate it is:

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

For more details see Sánchez (2014).

3.1.4 MINSADBAD Regression

This method consists in estimating the parameters β_0 and β_1 minimizing the sum of the absolute differences between absolute deviations. It is:

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

Where $d_i = Y_i - (\beta_0 + \beta_1 X_i)$. This method is easier to explain considering the linear multiple regression model. For more details it is recommended to read Sanjith & Elangovan (2014).

4 Conclusions

The way to address the teaching and learning processes has moved from a model focused on teaching to one focused on student learning and its relation to personal and contextual factors. This paradigm shift calls into question what is intended for the students to learn, what are the modalities and procedures most appropriate for them to develop and acquire these learnings and with which criteria and procedures are we going to check whether the student has finally acquired them and the transfer he can do. But it's also an invitation to support the professors who were mostly educated in the paradigm where the most important were the content and its teaching; therefore they require models and teaching strategies that facilitate the implementation of the teaching model focused on the student's learning.

Theory developed around the concept of regression (conventional view) can't be different and, therefore, it must be a teaching modality that complements the expositive method of teaching or master class, highlighting the previous methodological effort and/or parallel to the formal formulation of the concept regression, stimulating the discrimination based on the knowledge of non-conventional methods. It's considered a priority to drive studies that facilitate strategies for university professors to promote a new approach to the teaching and learning processes in the classroom. In this paper it's intended to facilitate some orientations that allow the concision of the teaching and learning processes of the regression, taking into account the methodological implications of this paradigm shift. This paper introduced a compilation of some non-conventional methods of linear regression, based in the L_1 norm, which give rise to estimators that are an alternative for sets of data that don't meet the assumptions of normality or errors and describe more clearly the reality of these as a didactical historical- evolutionary modality of teaching of the regression.

Received: January 12, 2016
Accepted: March 3, 2016

References

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.