



## A linear regression model for imprecise response

M.B. Ferraro<sup>a,\*</sup>, R. Coppi<sup>a</sup>, G. González Rodríguez<sup>b</sup>, A. Colubi<sup>c</sup>

<sup>a</sup> Dipartimento di Statistica, Probabilità e Statistiche Applicate, Sapienza Università di Roma, Italy

<sup>b</sup> European Centre for Soft Computing, Mieres, Spain

<sup>c</sup> Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, Spain

### ARTICLE INFO

#### Article history:

Received 25 March 2009

Received in revised form 23 March 2010

Accepted 4 April 2010

Available online 20 April 2010

#### Keywords:

Least-squares approach

Asymptotic distribution

LR fuzzy data

Interval data

Regression models

### ABSTRACT

A linear regression model with imprecise response and  $p$  real explanatory variables is analyzed. The imprecision of the response variable is functionally described by means of certain kinds of fuzzy sets, the *LR* fuzzy sets. The *LR* fuzzy random variables are introduced to model usual random experiments when the characteristic observed on each result can be described with fuzzy numbers of a particular class, determined by 3 random values: the center, the left spread and the right spread. In fact, these constitute a natural generalization of the interval data. To deal with the estimation problem the space of the *LR* fuzzy numbers is proved to be isometric to a closed and convex cone of  $\mathbb{R}^3$  with respect to a generalization of the most used metric for *LR* fuzzy numbers. The expression of the estimators in terms of moments is established, their limit distribution and asymptotic properties are analyzed and applied to the determination of confidence regions and hypothesis testing procedures. The results are illustrated by means of some case-studies.

© 2010 Elsevier Inc. All rights reserved.

## 1. Introduction

Different elements of a statistical problem may be imprecisely observed or defined. This has led to the development of various theories able to cope with an uncertainty which is not necessarily due to randomness: e.g. the methods based on imprecise probabilities (see, for instance, [37]), subjective probabilities (see, for instance, [34]) belief functions (see, for instance, [29]) or diverse approaches for fuzzy statistical analysis (see, for instance, [3,5] or [7]). In this paper we will consider a regression problem for a random experiment in which a fuzzy response and real-valued explanatory variables are observed. Actually, in many practical applications in public health, medical science, ecology, social or economic problems, many useful variables are vague, and the researchers find it easier to reflect the vagueness through fuzzy data than to discard it and obtain precise data. In addition, it is often less expensive to obtain an imprecise observation than to look for precise measurements of the variable of interest (see, for instance, [16]). Formally, any  $[0,1]$ -valued function determines a fuzzy set. However, in practice, the usual membership functions belong to some specific classes easier to fix and handle. In particular, the class of fuzzy numbers consisting of upper semi-continuous  $[0,1]$ -valued functions with compact support is rich enough to cover most of the applications (see, for instance [24] or [9]). However, this class is still very general, and many practitioners prefer to use simple shapes, as triangular or, slightly more general ones, as *LR*-fuzzy numbers, which are considered flexible enough to represent accurately their real-life data. For example, in agriculture quantitative soil data are unavailable over vast areas and imprecise measures, that can be modelled through *LR* fuzzy sets, are used (see [22]). Also in medical science symptoms,

\* Corresponding author.

E-mail addresses: [mariabrigida.ferraro@uniroma1.it](mailto:mariabrigida.ferraro@uniroma1.it) (M.B. Ferraro), [renato.coppi@uniroma1.it](mailto:renato.coppi@uniroma1.it) (R. Coppi), [gil.gonzalez@softcomputing.es](mailto:gil.gonzalez@softcomputing.es) (G. González Rodríguez), [colubi@uniovi.es](mailto:colubi@uniovi.es) (A. Colubi).

diagnosis and phenomena of disease may often lead to *LR* data (see, for instance, [6]). *LR*-type fuzzy data may also arise in other contexts, like image processing or artificial intelligence (see, for instance, [32,33]).

*LR* fuzzy sets are a generalization of intervals. Epidemiological research often entails the analysis of failure times subject to grouping, and the analysis with interval-grouped data is numerically simple and statistically meaningful (see [30,13,2]).

There are two different main lines concerning fuzzy regression problems in the literature; namely, the so-called fuzzy or possibilistic regression introduced by Tanaka [35] and widely analyzed since then (see, for instance, [36,25] and references therein) and the so-called least squares problems involving fuzzy data, also widely studied by Diamond [8], Näther [27], Krätschmer [20], González-Rodríguez et al. [15] and references therein. When the first one involves fuzzy data, an imprecise model focused on inclusion relations between actual and estimated outputs to explain the relationship is searched (see, for instance, [36]). In contrast, in the second approach, classical least squares fitting or statistical estimation problems for standard models involving fuzzy random variables are taken into consideration. Modelling this situation might be viewed as an extension of the classical error in variables models admitting measurement errors which are of nonrandom nature. From another point of view, this problem can be managed in a standard framework by means of an appropriate metric and through the concepts coherent with the space structure. Many of the above-mentioned fuzzy regression analyses have mainly focused on non probabilistic models. The only source of uncertainty accounted for in this case was the vagueness/imprecision of the data and/or of the regression parameters, and appropriate techniques of fuzzy/possibilistic analysis were utilized in this respect. Only a few papers have been devoted to regression methods able to cope with both imprecision and randomness (due to the data generation process). Among these we mention González-Rodríguez et al. [15], Näther [27] and Körner [18,19]. The present work is framed in the latter context. Specifically, a generalization of the work of Coppi et al. [4] is considered.

Coppi et al. [4] have proposed a linear regression model with crisp inputs and *LR* fuzzy response. The basic idea consists in modelling the centers of the response variable by means of a classical regression model, and simultaneously modelling the left and the right spread of the response through simple linear regressions on its estimated center. The study in Coppi et al. [4] is mainly descriptive, and the authors impose a non-negativity condition to the numerical minimization problem to avoid negative estimated spreads. In this work we propose an alternative model to overcome the non-negativity condition, because the inferences for models with non-negativity restrictions are more complex and less efficient (see, for instance, [23,12]).

The model may be looked at in the context of a multivariate regression problem. From a semantic viewpoint, it differs from the classical econometric models. In fact the equations related to the centers and the spreads jointly refer to a unique fuzzy variable, and therefore to a unique phenomenon which may be jointly affected by several (crisp) variables. Consequently, the approach we propose allows us to express and analyze the model within the context of classical multiple and multivariate regression models (see, for instance, [28]). Furthermore, unlike the model proposed by Coppi et al., in the proposed model the left and the right spread of the response are not modelled through simple linear regressions on its estimated center but by means of simple linear regressions on the explanatory variables. In this way, the parameter identification problem is avoided.

The rest of the paper is organized as follows. In Section 2 the way of modelling the imprecise response through *LR* fuzzy random variables is formalized. In Section 3 the variance of an *LR* fuzzy random variable is defined and some properties are proved. In Section 4 the new linear regression model is introduced, and the least squares estimators of the parameters are found and analyzed. Section 5 deals with asymptotic confidence regions and asymptotic hypothesis tests for the regression parameters. In Section 6 a real-life example with *LR* fuzzy data and another with interval data are illustrated. Finally, Section 7 contains some remarks and future directions.

## 2. Modelling the imprecise data

### 2.1. Fuzzy sets

A fuzzy set  $A$  of  $\mathbb{R}$  may be simply defined as a mapping  $A : \mathbb{R} \rightarrow [0, 1]$  verifying some conditions (see, for instance, Zadeh [39]). In practice there are some experiments whose results can be described by means of fuzzy sets of a particular class, determined by 3 values: the center, the left spread and the right spread. This type of fuzzy datum is called *LR fuzzy number* and is defined in the following way (see Fig. 1)

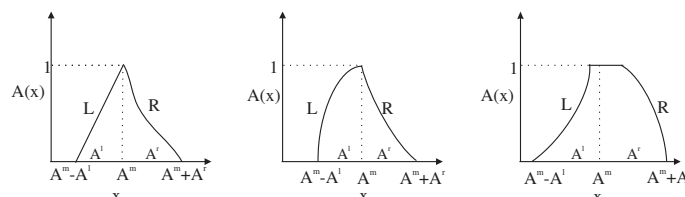


Fig. 1. Examples of *LR* membership functions.

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات