An Integrated approach for fuzzy logistic regression

Nimet Yapıcı Pehlivan
Department of Statistics, Faculty of Statistics,
University of Selçuk, 42075, Konya, Turkey
nimet@selcuk.edu.tr
0000-0002-7094-8097

Aynur Şahin
Department of Statistics, Faculty of Statistics,
University of Selçuk, 42075, Konya, Turkey
aynursahin@selcuk.edu.tr
0000-0003-1681-9298

Abstract

The aim of this study is to introduced an integrated fuzzy logistic regression approach to describe the relationship between crisp inputs and fuzzy binary output. For this reason, we integrated the fuzzy logistic regression methods proposed by Pourahmad et al. [17] and Sohn et al. [24] to define a possibility measure for each case and then used the logarithmic transformation of possibilistic odds as fuzzy output observations. To estimate the parameters of the fuzzy logistic regression model, Diamond [5]'s Fuzzy Least Squares (FLS) approach is used. A numerical example is presented and obtained results are compared with classic logistic regression model.

Keywords**:** Fuzzy Logistic Regression, Possibilistic odds, Fuzzy Least Squares.

Öz

Bulanık Lojistik Regresyon için Entegre Bir Yaklaşım

Bu çalışmanın amacı, kesin girdiler ile bulanık ikili çıktı arasındaki ilişkiyi tanımlamak için birleştirilmiş bulanık lojistik regresyon yaklaşımını tanıtmaktır. Bu nedenle, her bir durum için bir olasılık ölçüsü tanımlamak ve daha sonra olabilirlik oranlarının logaritmik dönüşümünü bulanık çıktı gözlemleri olarak kullanmak amacıyla Pourahmad ve ark. [17] ve Sohn ve ark. [24] tarafından önerilen bulanık lojistik regresyon yöntemleri entegre edilmiştir. Bulanık lojistik regresyon modelinin parametrelerini tahmin etmek için, Diamond [5]'ın Bulanık En Küçük Kareler (FLS) yaklaşımı kullanılmıştır. Sayısal bir örnek sunulmuş ve elde edilen sonuçlar klasik lojistik regresyon modeli ile karşılaştırılmıştır.

Anahtar sözcükler: Bulanık Lojistik Regresyon, Olabilirlik Oranları, Bulanık En Küçük Kareler.

1. Introduction

Regression analysis is one of the most widely used and powerful methodology to model the functional relationship between a response variable (dependent variable, output) and one or more explanatory variables (independent variables, inputs). In many real world problems, some observations couldn't be determined exactly and they need to be defined by linguistic terms or fuzzy numbers. Fuzzy regression (FR) analysis first introduced by Tanaka et al. [26], is an extension of classical regression analysis in which some elements of the model are represented by any type of fuzzy numbers. On the contrary to

classical regression that is based on probability theory, fuzzy regression could be based on possibility theory and fuzzy set theory. There are two main approaches for Fuzzy Regression: (i) Possibilistic Regression approach proposed by Tanaka et al. [26] which is based on possibility concepts and minimizes total vagueness of the estimated values for the outputs (see, Chang and Lee [3], Hojati et al. [6], Modarres et al.[10], Nasrabadi et al. [13], Özelkan and Duckstein [14], Peters [15], Sakawa and Yano [21], Tanaka and Watada [27], Tanaka [25], etc). (ii) Fuzzy Least Squares Method proposed by Diamond [5] and Celmins [2] simultaneously which is based on minimizing the errors between observed and estimated outputs (see, (D'urso [4], Diamond [5], Kao and Chyu [7], Khan and Valeo [8], Ming et al. [9], Yang and Lin [28], etc.).

Some studies on Fuzzy Regression models considered crisp (i.e. nonfuzzy) inputs, fuzzy parameters and fuzzy outputs (see, Hojati et al. [6], Nasrabadi et al. [13], Peters [15], Savic and Pedrycz [22], Shakouri and Nadimi [23], Tanaka [25], Tanaka and Watada [27], Tanaka et al. [26]). Some of them handled fuzzy inputs, crisp parameters and fuzzy outputs (see, Diamond [5], Kao and Chyu [7], Ming et al. [9]), while the other studies took into account fuzzy inputs, fuzzy outputs and also fuzzy parameters (see, Hojati et al. [6], Khan and Valeo [8], Sakawa and Yano [20], Sakawa and Yano [21], Yang and Lin [28]).

Logistic regression analysis have been used to model functional relationship between a binary response variable and one or more explanatory variables. Fuzzy Logistic Regression defined the relationship between crisp/fuzzy inputs and fuzzy binary outputs considering fuzzy regression approaches. Contrary to fuzzy regression models, there have been a limited number of studies on fuzzy logistic regression in the literature. To estimate parameters of the fuzzy logistic regression model, Tanaka et al. [26]'s possibilistic approach and Diamond [5]'s Fuzzy Least Squares method have been used by several authors: Pourahmad et al. [16] proposed fuzzy logistic regression model for crisp inputs and binary output which its value takes a number between 0 and 1. They used Tanaka et al. [26]'s possibilistic approach by introducing the concept of possibilistic odds to estimate the model parameters and also proposed some goodness-of-fit criteria. Pourahmad [17] presented Fuzzy Logistic Regression to model fuzzy binary outputs. Due to the vague nature of binary outputs, a fuzzy model based on possibility of success defined by linguistic terms was constructed. The logarithm transformation of possibilistic odds was modelled based on crisp inputs by using Extension Principle and Diamond [5]'s FLS method was used to estimate model parameters. The capability index was calculated for evaluating the model. Namdari et al.[11] aimed to assess the effect of folic acid on the appetite of preschool children 3 to 6 years old. Since appetite was measured by linguistic terms, a fuzzy logistic regression was applied by using Diamond [5]'s FLS method and the obtained results were compared with a ordinal logistic regression model. Namdari et al. [12] investigated fuzzy logistic regression model for crisp input and fuzzy binary output which was measured by linguistic terms. They developed least absolute deviations (LAD) method for modeling and compared the results with the least squares estimation (LSE) method by providing two estimation methods, min–max method and fitting method. Also, new goodness-of-fit indices, measure of performance based on fuzzy distance (Mp) and index of sensitivity (IS), were presented. Sohn et al. [24] presented a new technology credit scoring model based on the fuzzy logistic regression method with fuzzy input and fuzzy binary output data for predicting the possibility of loan default of firms. The estimation method of coefficients was conducted based on Yoon and Choi [29]'s FLS.

The aim of this study is to introduce an integrated approach for fuzzy logistic regression model with crisp inputs and fuzzy output based on Pourahmad et al. [17] and Sohn et al. [24]. To estimate the parameters of the fuzzy logistic regression model, FLS approach of Diamond [5] is used. A numerical example is presented and obtained results are compared with classical logistic regression model.

The remainder of this study is as follows: some basic definitions of fuzzy sets are given in Section 2. A brief review on logistic regression and detailed definition of fuzzy logistic regression methods are given in Section 3. In Section 4, an application of Pima India Diabetes data set is employed to investigate the proposed fuzzy logistic regression model and to compare with classical logistic regression. Finally, conclusions are drawn in Section 5.

2. Basic Definitons of Fuzzy Sets

This section gives basic definitions and concepts of fuzzy sets and arithmetic operations on fuzzy sets are given as [19]

Definition 2.1. A fuzzy set $\tilde{A}$ defined in space X is a set of pairs:

$$\tilde{A} = \left\{ (x, \mu(x)), x \in X \right\} \tag{2.1}$$

where: fuzzy set $\tilde{A}$ is characterized by its membership function $\mu_A(x): X \to [0,1]$.

Definition 2.2. α-level set of a fuzzy set $\tilde{A}$ in space X is defined as

$$\mu_{\tilde{A}}(x) \geq \alpha, \quad \alpha \in [0,1] \tag{2.2}$$

Definition 2.3. (Extension Principle) Let X be the Cartesian product of universes $(X_1, X_2, ..., X_n)$ and $(\tilde{A}, \tilde{A}, ..., \tilde{A})$ be fuzzy sets in $(X_1, X_2, ..., X_n)$, respectively. Suppose that f is a mapping from X to a universe Y, $y = f(x, x, ..., x) \in Y$. A fuzzy set $\tilde{B}$ in Y is defined by Extension principle as follows:

$$\tilde{B} = \begin{cases} \sup \min \left\{ \mu(x), \mu(x), ..., \mu(x) \right\}, \ f(y) \neq \phi \\ 0, \ otherwise \end{cases} \tag{2.3}$$

Definition 2.4. A triangular fuzzy number $\tilde{A} = (a_1, a_2, a_3)$ is characterized by following membership function,

$$\mu_{\tilde{A}}(x) = \begin{cases} 0, x < a_1, x > a_3 \\ \dfrac{x - a_1}{a_2 - a_1}, \quad a_1 \leq x \leq a_2 \\ \dfrac{a_3 - x}{a_3 - a_2}, \quad a_2 \leq x \leq a_3 \end{cases} \tag{2.4}$$

Definition 2.5. $\tilde{A} = (a_1, a_2, a_3)$ and $\tilde{B} = (b_1, b_2, b_3)$ are two triangular fuzzy numbers. Arithmetic operations on fuzzy numbers are considered as follows:

Addition: $\tilde{A}(+)\tilde{B} = (a_1 + b_1, a_2 + b_2, a_3 + b_3)$ $\tag{2.5}$

Substraction: $\tilde{A}(-)\tilde{B} = (a_1 - b_3, a_2 - b_2, a_3 - b_1)$ $\tag{2.6}$

Multiplication: $\tilde{A}(\times)\tilde{B} = (a_1 b_1, a_2 b_2, a_3 b_3)$ $\tag{2.7}$

Division: $\tilde{A}(\div)\tilde{B} = (\dfrac{a_1}{b_1}, \dfrac{a_2}{b_2}, \dfrac{a_3}{b_3})$ $\tag{2.8}$

Multiplication by constant k: $k(\times)\tilde{A} = \begin{cases} (ka_1, ka_2, ka_3), k > 0 \\ (ka_3, ka_2, ka_1), k < 0 \end{cases}$ (2.9)

Inverse: $\left(\tilde{A}\right)^{-1} = \left(\dfrac{1}{a_3}, \dfrac{1}{a_2}, \dfrac{1}{a_1}\right)$ (2.10)

3. Fuzzy Logistic Regression

This section presents a brief review on logistic regression and detailed definition of fuzzy logistic regression methods.

3.1. Logistic Regression

Logistic regression analysis have been used to model functional relationship between a binary dependent variable and one or more independent variables and frequently applied in clinical research, credit scoring, genetics, etc. Independent variables may be continuous, discrete, and binary or a mix of any of these. No assumptions are made about the distribution of these variables. Binary dependent variable with two categories $Y = 0,1$ (i.e. No/Yes or Failure/Success) has Bernoulli probability distribution. It can take the value of 1 with a probability of success (π), or the value of 0 with probability of failure ($1-\pi$). Due to the relationship between the independent and dependent variables is not linear, the logit transformation of π is used as follows:

$$\pi = \frac{1}{1 + e^{b_0 + b_1 x_1 + ... + b_n x_n}}$$ (3.1)

So, the form of the model in Eq. (3.1) is considered as:

$$\ln\left(\frac{\pi}{1-\pi}\right) = b_0 + b_1 x_1 + ... + b_n x_n$$ (3.2)

where: $\dfrac{\pi}{1-\pi} = e^{b_0 + b_1 x_1 + ... + b_n x_n}$ is called as "probability odds" and $b, j = 0,1,...,n$ indicates the model parameters [17].

3.2. Fuzzy Logistic Regression

Fuzzy Logistic Regression has defined the relationship between crisp/fuzzy independent variables and fuzzy binary dependent variables considering fuzzy regression approaches. To estimate parameters of the fuzzy logistic regression model, Tanaka et al. [26]'s possibilistic approach and Diamond [5]'s Fuzzy Least Squares method have been used. In the literature, contrary to fuzzy linear regression models, there have been only a few studies on fuzzy logistic regression, such as Pourahmad et al.[16, 17], Namdari et al.[11, 12], Sohn et al.[24].

Generally, when establishing a model for logistic regression, uncertain or fuzzy observations are eliminated. However, if there are too many uncertain or fuzzy observations, a model cannot be constructed since the probabilistic odds cannot be computed. Therefore, the concept of "possibilistic odds" can be considered for such observations. Suppose that $\tilde{Y}_i, i = 1, 2, ..., m$ is a fuzzy binary dependent variable and

$x_{ij}, i = 1, 2, ..., m; j = 1, 2, ..., n$ is a set of crisp independent variables, the definition of "possibilistic odds" given in Pourahmad et al.[17] is as follows:

$\mu_i, i = 1, 2, ..., m$ is the possibility of success, i.e. $\mu_i = Poss(Y_i \approx 1)$ and it could be defined by i) a real crisp value, $\mu \in \Re,\ 0 \le \mu \le 1$ or ii) a linguistic term, such as $\tilde{\mu}_i \in \{$Very low, Low, Medium, High, Very High$\}$. The ratio $\dfrac{\tilde{\mu}_i}{1 - \tilde{\mu}_i}, i = 1, 2, ..., m$ is handled as "possibility odds" for the observation i, which determines the possibility of success relative to the possibility of non-success.

In this study, we considered linguistic terms $\tilde{\mu} \in \{$Approximately 0, Approximately 1$\}$ which are characterized by triangular fuzzy numbers as Approximately 0=(0.01, 0.03, 0.05) and Approximately 1=(0.95, 0.97, 0.99) for the possibility of success as shown in Figure 1. Due to the computational simplicity in arithmetic operations, the triangular fuzzy numbers are the most widely used fuzzy number type in the literature and have a common range of uses.
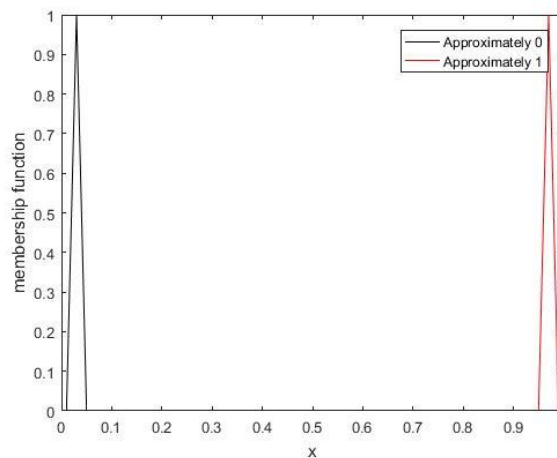

Figure 1. Membership functions of linguistic terms of Approximately 0 and Approximately 1

In the integrated method, the triangular fuzzy numbers $\tilde{Y}$ corresponding to $\tilde{\mu}$ are computed as $\tilde{Y} = \ln\left(\dfrac{\tilde{\mu}}{1 - \tilde{\mu}}\right)$ according to the Sohn et al. [24]. Then, a fuzzy logistic regression model with a fuzzy binary dependent variable proposed by Pourahmad et al. [17] is applied as follows:

$$\tilde{W}_i = \ln\left(\frac{\tilde{\mu}_i}{1 - \tilde{\mu}_i}\right) = \tilde{A}_0 + \tilde{A}_1 x_{i1} + ... + \tilde{A}_n x_{in}; i = 1, 2, ..., m \tag{3.3}$$

where; $\tilde{A}_j = (a_j, s_j)_T, j = 0, 1, ..., n$ indicates fuzzy parameters which are symmetrical triangular fuzzy numbers with center a_j and spread s_j. $\tilde{W}_i$ is the estimator of the logarithmic transformation of possibilistic odds and it is called as estimated fuzzy dependent variable. $\tilde{W}_i$ is a triangular fuzzy number denoted by $\tilde{W}_i = (f_i(a), f_i(s))_T, i = 1, 2, ..., m$ in which $f_i(a) = a_0 + a_1 x_{i1} + ... + a_n x_{in}$ and $f_i(s) = s_0 + s_1 x_{i1} + ... + s_n x_{in}$.

α -level set of $\tilde{W}_i$ is computed as:

$$\left(\tilde{W}_i\right)_\alpha = \left[f_i(a) - (1 - \alpha) f_i(s), f_i(a) + (1 - \alpha) f_i(s)\right] \tag{3.4}$$

In the method, an expert assigns a linguistic term $\tilde{\mu}_i$ as the possibility of success. Then, the logarithmic transformation of possibility odds are considered as observed dependent variables, $\tilde{w}_i$, as follows:

$$\tilde{w}_i = \ln\left(\frac{\tilde{\mu}_i}{1-\tilde{\mu}_i}\right), i = 1, 2, ..., m \tag{3.5}$$

The membership function of the $\tilde{w}_i$ is calculated from the defined membership function of $\tilde{\mu} = (l, m, r)$ by using Extension Principle. α-level set of $\tilde{w}_i$ is calculated based on $(\tilde{\mu})$ as:

$$(\tilde{w}_i)_\alpha = \left[\ln\left(\frac{l_{\tilde{\mu}_i}(\alpha)}{1-l_{\tilde{\mu}_i}(\alpha)}\right), \ln\left(\frac{r_{\tilde{\mu}_i}(\alpha)}{1-r_{\tilde{\mu}_i}(\alpha)}\right)\right] \tag{3.6}$$

where:

$$(\tilde{\mu}_i)_\alpha = \left[l_{\tilde{\mu}_i}(\alpha), r_{\tilde{\mu}_i}(\alpha)\right]$$

$$l_{\tilde{\mu}_i}(\alpha) = m_{\mu_i} - (1-\alpha)(m_{\mu_i} - l_{\mu_i})$$

$$r_{\tilde{\mu}_i}(\alpha) = m_{\mu_i} + (1-\alpha)(r_{\mu_i} - m_{\mu_i})$$

To obtain an optimal model for Eq (3.3), Sum of Squares Error (SSE) between $\tilde{w}$ and $\tilde{W}$, $i = 1, 2, ..., m$ is minimized by Diamond [5]'s FLS method by:

$$SSE = \sum_{i=1}^{m} d(\tilde{w}_i, \tilde{W}_i)^2 = \sum_{i=1}^{m} \int_0^1 \left[f(\alpha)d^2(\tilde{w}_i(\alpha), \tilde{W}_i(\alpha)d\alpha\right]^{1/2} \tag{3.7}$$

By using Eq (3.4) and Eq(3.6), SSE is computed by,

$$SSE = \sum_{i=1}^{m} \int_0^1 f(\alpha)\left[\left[\ln\left(\frac{l_{\tilde{\mu}_i}(\alpha)}{1-l_{\tilde{\mu}_i}(\alpha)}\right) - f_i(a) - (\alpha-1)f_i(s)\right]^2 + \left[\ln\left(\frac{r_{\tilde{\mu}_i}(\alpha)}{1-r_{\tilde{\mu}_i}(\alpha)}\right) - f_i(a) - (1-\alpha)f_i(s)\right]^2\right]d\alpha \tag{3.8}$$

By minimizing the SSE with respect to a_j and s_j, $j = 0, 1, ..., n$ following equations are obtained:

$$\frac{\partial SSE}{\partial a} = \sum\left(\int 2\alpha x\left[2f(a) - \ln(-) - \ln(-)\right]d\alpha\right) = 0 \tag{3.9}$$

$$\frac{\partial SSE}{\partial s_j} = \sum_{i=1}^{m}\left(\int_0^1 2\alpha(1-\alpha)x_{ij}\left[2(1-\alpha)f_i(s) + \ln\left(\frac{l_{\tilde{\mu}_i}(\alpha)}{1-l_{\tilde{\mu}_i}(\alpha)}\right) - \ln\left(\frac{r_{\tilde{\mu}_i}(\alpha)}{1-r_{\tilde{\mu}_i}(\alpha)}\right)\right]d\alpha\right) = 0 \tag{3.10}$$

The quantities in Eq (3.7) and Eq (3.8) depend on definition of possibility of success ($\tilde{\mu}_i$) assigned to each observation by the expert. Substituting these values and calculating related integrals, the following equations are derived:

$$a_0 \sum_{i=1}^{m} x_{i0} x_{ij} + a_1 \sum_{i=1}^{m} x_{i1} x_{ij} + ... + a_n \sum_{i=1}^{m} x_{in} x_{ij} = \sum_{i=1}^{m} z_i x_{ij}, \; j =, 1, ..., n \tag{3.11}$$

$$s_0 \sum_{i=1}^{m} x_{i0} x_{ij} + s_1 \sum_{i=1}^{m} x_{i1} x_{ij} + ... + s_n \sum_{i=1}^{m} x_{in} x_{ij} = \sum_{i=1}^{m} k_i x_{ij}, \; j =, 1, ..., n \tag{3.12}$$

Where: $x_{i0} = 1, i = 1, 2, ..., m$ and z_i, k_i are the results of integral computation for each observation. Eq (3.9) and Eq (3.10) can be rewritten in a matrix form $aA = Z$ and $sA = K$ in which

$$A = X^T X, \quad X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1n} \\ 1 & x_{21} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & \cdots & x_{mn} \end{bmatrix}_{m \times (n+1)} \tag{3.13}$$

$$a = (a, a, ..., a), \; Z = \left(\sum_{i=1}^{m} z_i x_{i0}, \sum_{i=1}^{m} z_i x_{i1}, ..., \sum_{i=1}^{m} z_i x_{in} \right)^T \tag{3.14}$$

$$s = (s_0, s_1, ..., s_n)^T, \; K = \left(\sum_{i=1}^{m} k_i x_{i0}, \sum_{i=1}^{m} k_i x_{i1}, ..., \sum_{i=1}^{m} k_i x_{in} \right)^T \tag{3.15}$$

Therefore, a and s are computed from following equations as in [18];

$$a = A^{-1} Z \tag{3.16}$$

$$s = A^{-1} K \tag{3.17}$$

Finally, estimated fuzzy logistic regression model is obtained as in [17],

$$\tilde{W}_i = (a_0, s_0) + (a_1, s_1) x_{i1} + ... + (a_n, s_n) x_{in}; i = 1, 2, ..., m \tag{3.18}$$

4. Application

In this study, we discussed fuzzy logistic regression model based on possibilistic odds and applied on a real dataset "Pima India Diabetes", where eight inputs variables and one output variable, taken from UCI Machine Learning Repository [18]. A part of the data set is shown in Table A.1. The output variable in this data set consists of two categories: "0" and "1". "0" indicates that a patient has not diabetes and "1" indicates that a patient has diabetes. However, sometimes physicians may not make a definite decision whether a patient has diabetes, and this patient can not be categorized. In short, there may be a fuzziness about the status of the patient and classical logistic regression model cannot be applied. In order to overcome this situation, fuzzy logistic regression model could be applied.

Inputs:

X1: Number of times pregnant

X2: Plasma glucose concentration a 2 hours in an oral glucose tolerance test

X3: Diastolic blood pressure (mm Hg)

X4: Triceps skin fold thickness (mm)

X5: 2-Hour serum insulin (mu U/ml)

X7: Diabetes pedigree function

X8: Age (years)

Binary Output:

Y: Class variable which takes values 0 or 1, where "0" denotes that the patient has not diabetes and "1" denotes that the patient has diabetes.

In this study, we considered the value $\tilde{\mu}$ of the fuzzy binary dependent variable for observation i which can take linguistic term $\tilde{\mu} \in \{$Approximately 0, Approximately 1$\}$ as the possibility of success. Since the binary dependent variable is fuzzy, the possibilistic odds is calculated. Definitions of linguistic terms $\tilde{\mu}_i$ can be given in the form of triangular fuzzy number based on its membership function. The triangular fuzzy numbers $\tilde{Y}$ corresponding to $\tilde{\mu}$ are computed as shown in Table according to Sohn et al. [24].

Table 1. Transformations of triangular fuzzy number ($\tilde{\mu}$) and corresponding $\tilde{Y}$

Crisp output	Linguistic terms of fuzzy binary dependent variable	$\tilde{\mu}$	$\tilde{Y} = \ln\left(\frac{\tilde{\mu}}{1-\tilde{\mu}}\right)$
0	Approximately 0	(0.01, 0.03, 0.05)	(-4.5951, -3.4761, -2.9444)
1	Approximately 1	(0.95, 0.97, 0.99)	(2.9444, 3.4761, 4.5951)

In this study, we construct the fuzzy logistic regression model to describe the relationship between eight crisp inputs and a fuzzy binary output. For this reason, we integrated the fuzzy logistic regression methods proposed by Pourahmad et al. [17] and Sohn et al. [24] to define a possibility measure for each case and then use the logarithmic transformation of possibilistic odds as fuzzy output observations. By this method, Diamond [5]'s fuzzy least squares method in fuzzy linear regression can be used to estimate model parameters.

For $i=1$ ($\tilde{\mu}_1 \in \{$Approximately 0$\}$), estimated $\tilde{W}_1$ is

$$\tilde{W}_1 = \ln\left(\frac{\tilde{\mu}_1}{1-\tilde{\mu}_1}\right) = A_0 + A_1 x_{11} + ... + A_8 x_{18} = \left(f_1(a), f_1(s)\right)_T$$

α -level set of $\tilde{W}_1$ is calculated by:

$$\left(\tilde{W}_1\right)_\alpha = \left[f_1(a) - (1-\alpha)f_1(s), f_1(a) + (1-\alpha)f_1(s)\right]$$

where;

$$f_1(a) = a_0 + a_1 x_{11} + ... + a_8 x_{18} = a_0 + 6a_1 + ... + 50a_8$$

$$f_1(s) = s_0 + s_1 x_{11} + ... + s_8 x_{18} = s_0 + 6s_1 + ... + 50s_8$$

α -level set of $\tilde{w}_1$ is computed as follows:

$$\left(\tilde{w}_1\right)_\alpha = \left[\ln\left(\frac{l_{\tilde{\mu}_1}(\alpha)}{1-l_{\tilde{\mu}_1}(\alpha)}\right), \ln\left(\frac{r_{\tilde{\mu}_1}(\alpha)}{1-r_{\tilde{\mu}_1}(\alpha)}\right) \right]$$

where;

$$\left(\tilde{\mu}_1\right)_\alpha = \left[l_{\tilde{\mu}_1}(\alpha), r_{\tilde{\mu}_1}(\alpha) \right] = \left[0.97 - 0.02(1-\alpha), 0.97 + 0.02(1-\alpha) \right]$$

The squares error between $\tilde{w}_1$ and $\tilde{W}_1$ for *i*=1 is considered as follows:

$$SE_1 = \int_0^1 f(\alpha)\left[\left[\ln\left(\frac{0.97-0.02(1-\alpha)}{1-0.97+0.02(1-\alpha)}\right) - f_1(a) - (\alpha-1)f_1(s) \right]^2 + \left[\ln\left(\frac{0.97+0.02(1-\alpha)}{1-0.97-0.02(1-\alpha)}\right) - f_1(a) - (1-\alpha)f_1(s) \right]^2 \right] d\alpha$$

For *i*=2 ($\tilde{\mu}_2 \in \{$Approximately 1$\}$), estimated $\tilde{W}_2$ is

$$\tilde{W}_2 = \ln\left(\frac{\tilde{\mu}_2}{1-\tilde{\mu}_2}\right) = A_0 + A_1 x_{21} + ... + A_8 x_{28} = \left(f_2(a), f_2(s) \right)_T$$

α -level set of $\tilde{W}_2$ is calculated by:

$$\left(\tilde{W}\right) = \left[f(a) - (1-\alpha)f(s), f(a) + (1-\alpha)f(s) \right]$$

where;

$$f_2(a) = a_0 + a_1 x_{21} + ... + a_8 x_{28} = a_0 + 1a_1 + ... + 31a_8$$

$$f_2(s) = s_0 + s_1 x_{21} + ... + s_8 x_{28} = s_0 + 1s_1 + ... + 31s_8$$

α -level set of $\tilde{w}_2$ is computed as follows:

$$\left(\tilde{w}_2\right)_\alpha = \left[\ln\left(\frac{l_{\tilde{\mu}_2}(\alpha)}{1-l_{\tilde{\mu}_2}(\alpha)}\right), \ln\left(\frac{r_{\tilde{\mu}_2}(\alpha)}{1-r_{\tilde{\mu}_2}(\alpha)}\right) \right]$$

where;

$$\left(\tilde{\mu}_2\right)_\alpha = \left[b_1, b_2 \right] = \left[0.03 - 0.02(1-\alpha), 0.03 + 0.02(1-\alpha) \right]$$

The squares error between $\tilde{w}_2$ and $\tilde{W}_2$ for $i=2$ is considered as follows:

$$SE_2 = \int_0^1 f(\alpha) \left[\left[\ln\left(\frac{0.03 - 0.02(1-\alpha)}{1 - 0.03 + 0.02(1-\alpha)} \right) - f_2(a) - (\alpha - 1)f_2(s) \right]^2 + \left[\ln\left(\frac{0.03 + 0.02(1-\alpha)}{1 - 0.03 - 0.02(1-\alpha)} \right) - f_2(a) - (1-\alpha)f_2(s) \right]^2 \right] d\alpha$$

After computing the SSE over all observations, SSE is minimized with respect to a_j and s by using Eqs (3.9)-(3.10). Then, a and s are computed as follows:

$$a = A^{-1}Z = \left(-9.5236, 0.1448, 0.0416, -0.0164, 0.0011, -0.0013, 0.0932, 1.0357, 0.0184 \right)$$

$$s = A^{-1}K = \left(0.7338, 0, 0, 0, 0, 0, 0, 0 \right)$$

where;

$$A = X\,X, \qquad X = \begin{bmatrix} 1 & 6 & \cdots & 50 \\ 1 & 1 & \cdots & 31 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 23 \end{bmatrix}_{768 \times 9}$$

$$Z = \left(\sum_{i=1}^{768} z_i x_{i0}, \sum_{i=1}^{768} z_i x_{i1}, ..., \sum_{i=1}^{768} z_i x_{i8} \right)^T$$
$$= \left(-815.9672, -1213.4, -60258.5, -53146.9, -13688.6, -26399.4, -20166.7, -236.817, -19910.3 \right)$$

$$K = \left(\sum_{i=1}^{768} k_i x_{i0}, \sum_{i=1}^{768} k_i x_{i1}, ..., \sum_{i=1}^{768} k_i x_{i8} \right)^T$$
$$= \left(563.5584, 2166.911, 68131.13, 38944.97, 11573.49, 44971.67, 18029.69, 265.9299, 18733.18 \right)$$

Finally, fuzzy logistic regression model is constructed as follows

$$\tilde{W}_i = \ln\left(\frac{\tilde{\mu}}{1-\tilde{\mu}} \right), i = 1, 2, ..., 768$$
$$= (-9.5235, 0.7338) + (0.1448, 0)x_{i1} + (0.0416, 0)x_{i2} + (-0.0164, 0)x_{i3} + (0.0011, 0)x_{i4} + (-0.0013, 0)x_{i5}$$
$$+ (0.0932, 0)x_{i6} + (1.0357, 0)x_{i7} + (0.0184, 0)x_{i8}$$

Table 2. Characteristics of Pima India diabetes dataset and estimated parameters of classical/fuzzy logistic regression $\tilde{Y}$

	Min	Max	Mean±SD	Classical Logistic Regression		Fuzzy Logistic Regression	
				b	Std.error	a	s
Intercept	-	-	-	-8.405	0.717	-9.5236	0.7338
X1	0.00	17.00	3.84±3.36	0.123	0.032	0.1448	0
X2	0.00	199.00	120.89±31.97	0.035	0.004	0.0416	0
X3	0.00	122.00	69.10±19.35	-0.013	0.005	-0.0164	0
X4	0.00	99.00	20.53±15.95	0.001	0.007	0.0011	0
X5	0.00	846.00	79.99±115.24	-0.001	0.001	-0.0013	0

X6	0.00	67.10	31.99±7.88	0.090	0.015	0.0932	0
X7	0.08	2.42	0.47±0.33	0.015	0.009	1.0357	0
X8	21.00	81.00	33.24±11.76	0.945	0.299	0.0184	0

We evaluated the proposed model in terms of the sensitivity, specificity and accuracy which are frequently used to define the prediction ability of the logistic regression model with a binary output (Table A.2). In order to compute the sensitivity, specificity and accuracy by using Eqs (A.1)-(A.3), the estimated fuzzy binary output is defuzzified by using Center of Gravity (CoA) method and converted into crisp ones for fuzzy logistic regression model. Classification table for classical and fuzzy logistic regression is shown in Table 3.

Table 3. Classification table for classical/fuzzy logistic regression

	Classical Logistic Regression				**Fuzzy Logistic Regression**		
	Observed				**Observed**		
Estimated	**1**	**0**	**Total**	**Estimated**	**1**	**0**	**Total**
1	156	55	211	**1**	154	53	207
0	112	445	557	**0**	114	447	561
Total	268	500	768	**Total**	268	500	768

Sensitivity, specificity, and accuracy are the terms that are generally associated with a binary classification test to statistically measure the performance of the test. In a binary classification, a data set is divided into two categories based on whether they have common properties or not by identifying their significance. "Sensitivity" indicates, how well the test predicts one category and "specificity" measures how well the test predicts the other category, wheares "accuracy" is expected to measure how well the test predicts both categories [1].

According to the Table 3, sensitivity, specificity and accuracy are computed as 0.8940, 0.5746 and 0.7825 for fuzzy logistic regression, while 0.8900, 0.5820 and 0.7825 for classical logistic regression. Sensitivity of the fuzzy logistic regression model is higher than that of the classical logistic regression model, whereas specifity is lower. On the other hand, accuracy is same for both method. It can be concluded that the prediction performance of fuzzy logistic regression is similar to the classical logistic regression. However, even a small increasing in sensitivity may be very important to us because the sensitivity shows how the test detects a disease. For this reason, we can conclude that the detection performance of the fuzzy logistic regression is relatively better than the classical logistic regression.

5. Conclusion

In this study, we introduced fuzzy logistic regression model with crisp dependent and fuzzy independent variables based on possibilistic odds presented by Pourahmad et al.[16, 17] and Sohn et al.[24]. We used FLS approach of Diamond [5] to estimate the parameters of the fuzzy logistic regression model. A numerical example conducted on Pima India Diabetes data set which have eight crisp input variables and one fuzzy binary output variable, is presented. Estimated parameters are obtained from both classical and fuzzy logistic regression models and classification table is set. Utilizing the classification table, we compared both methods in terms of the sensitivity, specificity and accuracy which are frequently used to define the prediction ability of the logistic regression model with a binary output. It is concluded that the detection performance of fuzzy logistic regression model is higher than classical logistic regression model according to sensitivity.

For future research, fuzzy logistic regression model based on Diamond [5]'s fuzzy least squares approach could be used to handle both fuzzy input and fuzzy output. Further, fuzzy logistic regression model based on Tanaka et al. [26]'s possibilistic approach could be used for data sets which have crisp or fuzzy input

and fuzzy output. Moreover, both possibilistic and least squares approaches for fuzzy logistic regression model should be applied and tested on complete real life problems.

References

[1] B.L. Aswathi, 2009, Sensitivity, Specificity, Accuracy and the relationship between them, in: *Bioinformatics*.

[2] A. Celmiņš, 1987, Least squares model fitting to fuzzy vector data, *Fuzzy Sets and Systems*, 22, 245-269.

[3] P.T. Chang, E.S. Lee, 1994, Fuzzy linear regression with spreads unrestricted in sign, *Computers & Mathematics with Applications*, 28, 61-70.

[4] P. D'Urso, 2003, Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data, *Computational Statistics & Data Analysis*, 42, 47-72.

[5] P. Diamond, 1988, Fuzzy least squares, *Information Sciences*, 46, 141-157.

[6] M. Hojati, C.R. Bector, K. Smimou, A simple method for computation of fuzzy linear regression, *European Journal of Operational Research*, 166 (2005) 172-184.

[7] C. Kao, C.-L. Chyu, 2002, A fuzzy linear regression model with better explanatory power, *Fuzzy Sets and Systems*, 126, 401-409.

[8] U.T. Khan, C. Valeo, 2015, A new fuzzy linear regression approach for dissolved oxygen prediction, *Hydrological Sciences Journal*, 60, 1096-1119.

[9] M. Ming, M. Friedman, A. Kandel, 1997, General fuzzy least squares, *Fuzzy Sets and Systems*, 88, 107-118.

[10] M. Modarres, E. Nasrabadi, M.M. Nasrabadi, 2005, Fuzzy linear regression models with least square errors, *Applied Mathematics and Computation*, 163, 977-989.

[11] M. Namdari, A. Abadi, S.M. Taheri, M. Rezaei, N. Kalantari, N. Omidvar, 2014, Effect of folic acid on appetite in children: Ordinal logistic and fuzzy logistic regressions, *Nutrition*, 30, 274-278.

[12] M. Namdari, J.H. Yoon, A. Abadi, S.M. Taheri, S.H. Choi, 2015, Fuzzy logistic regression with least absolute deviations estimators, *Soft Computing*, 19, 909-917.

[13] M.M. Nasrabadi, E. Nasrabadi, A.R. Nasrabady, 2005, Fuzzy linear regression analysis: a multi-objective programming approach, *Applied Mathematics and Computation*, 163, 245-251.

[14] E.C. Özelkan, L. Duckstein, 2000, Multi-objective fuzzy regression: a general framework, *Computers & Operations Research*, 27, 635-652.

[15] G. Peters, 1994, Fuzzy linear regression with fuzzy intervals, *Fuzzy Sets and Systems*, 63, 45-55.

[16] S. Pourahmad, S. Ayatollahi, S.M. Taheri, 2011a, Fuzzy Logistic Regression: A New Possibilistic Model And Its Application In Clinical Vague Status, *Iranian Journal of Fuzzy Systems*, 8, 1-17.

[17] S. Pourahmad, S.M.T. Ayatollahi, S.M. Taheri, Z.H. Agahi, 2011b, Fuzzy logistic regression based on the least squares approach with application in clinical studies, *Computers & Mathematics with Applications*, 62, 3353-3365.

[18] UCI Machine Learning Repository, Pima India Data Sets. http://archive.ics.uci.edu/ml/datasets/Pima+%20Indians+Diabetes.

[19] T.J. Ross, 2010, Fuzzy Logic with Engineering Applications, John Wiley & Sons, Ltd.

[20] M. Sakawa, H. Yano, Multiobjective fuzzy linear regression analysis for fuzzy input-output data, *Fuzzy Sets and Systems*, 47 (1992) 173-181.

[21] M. Sakawa, H. Yano 1992a, Fuzzy linear regression analysis for fuzzy input-output data, *Information Sciences*, 63, 191-206.

[22] D.A. Savic, W. Pedrycz, 1991, Evaluation of fuzzy linear regression models, *Fuzzy Sets and Systems*, 39, 51-63.

[23] H. Shakouri G, R. Nadimi, 2009, A novel fuzzy linear regression model based on a non-equality possibility index and optimum uncertainty, *Applied Soft Computing*, 9, 590-598.

[24] S.Y. Sohn, D.H. Kim, J.H. Yoon, 2016, Technology credit scoring model with fuzzy logistic regression, *Applied Soft Computing*, 43, 150-158.

[25] H. Tanaka, 1987, Fuzzy data analysis by possibilistic linear models, *Fuzzy Sets and Systems*, 24, 363-375.

[26] H. Tanaka, S. Uejima, K. Asai, 1982, Linear Regression Analysis with Fuzzy Model, *IEEE Transactions on Systems, Man, and Cybernetics*, 12, 903-907.

[27] H. Tanaka, J. Watada, 1988, Possibilistic linear systems and their application to the linear regression model, *Fuzzy Sets and Systems*, 27, 275-289.

[28] M.-S. Yang, T.-S. Lin, 2002, Fuzzy least-squares linear regression analysis for fuzzy input–output data, *Fuzzy Sets and Systems*, 126, 389-399.

[29] J.H. Yoon, S.H. Choi, 2013, Fuzzy Least Squares Estimation with New Fuzzy Operations, in: R. Kruse, R.M. Berthold, C. Moewes, Á.M. Gil, P. Grzegorzewski, O. Hryniewicz (Eds.) Synergies of Soft Computing and Statistics for Intelligent Data Analysis, Springer Berlin Heidelberg, Berlin, Heidelberg, s. 193-202.

APPENDIX

Table A.1. A part of Pima India diabetes dataset

Observation number (*i*)	X1	X2	X3	X4	X5	X6	X7	X8	Linguistic terms of Fuzzy binary dependent variable (Y)
1	6	148	72	35	0	33.6	0.6270	50	Approximately 1
2	1	85	66	29	0	26.6	0.3510	31	Approximately 0
3	8	183	64	0	0	23.3	0.6720	32	Approximately 1
4	1	89	66	23	94	28.1	0.1670	21	Approximately 0
5	0	137	40	35	168	43.1	2.2880	33	Approximately 1
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
766	5	121	72	23	112	26.2	0.2450	30	Approximately 0
767	1	126	60	0	0	30.1	0.3490	47	Approximately 1
768	1	93	70	31	0	30.4	0.3150	23	Approximately 0

Table A.2. The formulas of Sensitivity, Specificity and Accuracy [24]

	Classical / Fuzzy Logistic Regression		
	Observed		
Estimated	**Pozitive**	**Negative**	**Total**
Pozitive	True Pozitive (TP)	False Pozitive (FP)	TP+FP
Negative	False Negative (FN)	True Negative (TN)	FN+TN
Total	TP+FN	FP+TN	TP+FN+FP+TN

Sensitivity$=1-\{FP/(FP+TN)\}$ \hfill (A.1)

Specificity$=1-\{FN/(FN+TP)\}$ \hfill (A.2)

Accuracy$=\{(TP+TN)/(TP+FN+FP+TN)\}$ \hfill (A.3)