Fuzzy Logistic Regression Using Bootstrap

Farzane Moradi[1], Ali Shadrokh[2], Alireza Arabpour[3]

1. Payame Noor University Tehran, Iran
2. Payame Noor University Tehran, Iran
3. Shahid Bahonar University Kerman, Iran.

Corresponding Author: Farzane Moradi

Abstaract: In this paper, we propose an adaptive technique for the prediction of binary response variable by combining the fuzzy least squares method with statistical logistic regression. Also for each α-cut, using bootstrap technique, we discuss the problem of statistical inference when we have crisp inputs and fuzzy outputs. A simulation study and numerical example in the clinical field are provided to check the efficiency of the proposed approach.
Key words: Fuzzy logistic regression, least squares method, bootstrap, parameters.

INTRODUCTION

We often use regression analysis to model the relationship between dependent (response) and independent (explanatory) variables. In traditional regression analysis,residuals are assumed to be due to random errors. However, the residual errors aresometimes due to the indefiniteness of the model structure or imprecise observations. Theuncertainty in this type of regression model is not random, but fuzzy. Since Zadeh (1965) introduced the concept of fuzzy sets, fuzziness has received increasing attention and fuzzy data analysis has become increasingly important.

There are two categories of fuzzy regression analysis.

Farzane Moradi is a Ph.D. student in Department of Statistics Payame Noor University of Tehran, Iran.Alireza Arabpour is an Assistant Professor at Department of Statistics, Faculty of Mathematics and Computer Shahid Bahonar University of Kerman, Iran. Ali Shadrokh is an Associate Professor at Department of Statistics, Payame Noor University of Tehran, Iran.

The first is the possibilistic method which is based on possibility concepts. Possibilistic regression analysis uses a fuzzy linear system as the regression model and the total vagueness of the estimated values for the dependent variables is minimized. This analysis was first proposed by Tanaka et al. (1982).

The second category of fuzzy regression analysis adopts the fuzzy least squares method (FLSM) for minimizing errors between the observed and estimated outputs. This approach was introduced and developed by Celmins and Diamond. The advantage of the possibilistic model is in its simplicity in programming and computation, and that of the fuzzy least squares method in obtaining the minimum degree of fuzziness between the observed and estimated outputs.

The logistic regression is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is a binary variable (which takes one of two possible values). However, due to the vague nature of some binary observations, no probability distribution can be considered for these data and the ordinary logistic regression may not be appropriate. For example, consider observations that are described only in linguistic terms (such as fair, good, and excellent). Fuzzy set theory provides a means for modeling such linguistic variables utilizing fuzzy membership functions and hence fuzzy logistic regression was proposed to deal with such data sets. In contrast to the ordinary logistic regression that is based on probability theory, fuzzy logistic regression can be based on possibility theory and fuzzy set theory.

In this paper, we perform logistic regression analysis with fuzzy data by using bootstrap techniques. It is well-known that the bootstrap procedure converges swiftly and bootstrapping is a general approach to statistical inference based on building a sampling distribution for a statistic by resampling from the data at hand. We shall also discuss estimation and hypothesis testing for the parameter of a fuzzy logistic regression model.

Preliminaries

A fuzzy subset of an universe set X is specified by a membership function $\mu_A: X \to [0,1]$. The collection of all the fuzzy subsets of X is denoted by $F(X)$.
If f is a mapping from X to a universe Y $(f: X \to Y)$ and $A \in F(X)$ then the extension principle allows us to define a fuzzy set $B = f(A)$ in Y with following membership function:

$$\mu_B(y) = \begin{cases} \sup_{x \in f^{-1}(y)} \mu_A(x) & \text{if } f^{-1}(y) \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

Definition 1**:** A fuzzy subset A of the set of real numbers $\mathbb{R}$ with membership function $\mu_A: \mathbb{R} \to [0,1]$ is called a fuzzy number if
(i) A is normal, i.e. there exists an element z_0 such that $\mu_A(z_0) = 1$;
(ii) A is convex, i.e. $\forall z_1, z_2 \in \mathbb{R}$ $\mu_A(\lambda z_1 + (1-\lambda) z_2) \geq \mu_A(z_1) \wedge \mu_A(z_2), \forall \lambda \in [0,1]$;
(iii) μ_A is upper semi-continuous;
(iv) $\mathrm{supp}(A) = \overline{\{z \in \mathbb{R} : \mu_A(z) > 0\}}$ is bounded.
Definition 2**:** A fuzzy number A can be represented as a family of sets called α-cuts, A_α, defined as:
$A_\alpha = \{z \in \mathbb{R} : \mu_A(z) \geq \alpha\}$ for $0 < \alpha \leq 1$
and
$A_0 = \overline{\{z \in \mathbb{R} : \mu_A(z) > 0\}}$ for $\alpha = 0$.
Based on the resolution identity we get $A = \bigcup_{\alpha \in [0,1]} \alpha A_\alpha$.
Definition 3**:** The arithmetic operations for two fuzzy numbers A and B with α-cuts $A_\alpha = [A_\alpha^L, A_\alpha^U]$ and $B_\alpha = [B_\alpha^L, B_\alpha^U]$ are defined as follow:
$A_\alpha + B_\alpha = [A_\alpha^L + B_\alpha^L, A_\alpha^U + B_\alpha^U]$.
For given $k \in \mathbb{R}$:

$$k.A_\alpha = \begin{cases} [kA_\alpha^L, kA_\alpha^U] & \text{if } k \geq 0 \\ [kA_\alpha^U, kA_\alpha^L] & \text{if } k < 0 \end{cases}$$

$k + A_\alpha = [k + A_\alpha^L, k + A_\alpha^U]$.
For subtraction we use the general Hukuhara difference [13]:
$[A_\alpha^L, A_\alpha^U] \ominus_g [B_\alpha^L, B_\alpha^U] = [C_\alpha^-, C_\alpha^+]$
where $C_\alpha^- = \min\{A_\alpha^L - B_\alpha^L, A_\alpha^U - B_\alpha^U\}$ and $C_\alpha^+ = \max\{A_\alpha^L - B_\alpha^L, A_\alpha^U - B_\alpha^U\}$.
We can also represent an α-cut, A_α, by its midpoint and width, i.e. $A_\alpha = (A_\alpha^C, A_\alpha^W)$ where $A_\alpha^C = \frac{1}{2}(A_\alpha^L + A_\alpha^U)$ and $A_\alpha^W = \frac{1}{2}(A_\alpha^U - A_\alpha^L)$.
In this paper we define the Euclidean distance between α-cuts of two fuzzy numbers A and B as:
$d(A_\alpha, B_\alpha) = \sqrt{(A_\alpha^C - B_\alpha^C)^2 + (A_\alpha^W - B_\alpha^W)^2}$. (1)
Ordinary Logistic Regression

In ordinary logistic regression we have a binary dependent variable (that only contains data coded as 1 for success and 0 for failure) and one or more independent variables. The distribution of such a dependent variable is Bernoulli with success probability p. Now, we need to define a link function. We know the identity link is not suitable because we have the problem of non-linearity. There are many different link functions, but the best (or the easiest to interpret) is the logit function. The logit function is the logarithm of the probability odds. Probability odds are defined as the ratio of the probability of success and the probability of failure $\left(\frac{p}{1-p}\right)$ and range between 0 and infinity. So the logit transformation is defined as $\mathrm{logit}(p) = \log \frac{p}{1-p}$. This function spreads the probabilities over the real number line ($\mathbb{R}$), and hence is a suitable link function. Our logistic regression model may now be written as follows:
$\log \frac{p_i}{1-p_i} = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i$, $i = 1, \ldots, n$.

In such models the parameters can be estimated by maximum likelihood method. The approach used for testing the significance of the coefficients in logistic regression is very similar to the one in linear regression, however it uses the likelihood function for a binary outcome variable. The confidence interval estimators for coefficients are obtained based on their respective tests.

Fuzzy Logistic Regression

A solution to a nonlinear regression problem, especially the logistic regression, uses the adaptive method. In this method by transforming the variables, the relationship between them becomes linear. In the logistic case the transformation used for linearity is the logit function. So in the adaptive model the logit transformation of the output observation is linearly related to X_i's.

Studies on the applications of an adaptive fuzzy regression model have been published; Dom et al. and Nagar and Srivastava simultaneously used the adaptive technique in the prediction of a binary response variable . Their model expresses the fuzzy relation between crisp inputs and crisp output observations. They tested their model on an oral cancer dataset and compared it with the fuzzy neural network method. Pourahmad et al. applied the adaptive model using the possibilistic approach and least squares method. In the first case they applied possibility concepts in fuzzy logistic regression model introducing the term of "possibilistic odds". Their model with crisp input, crisp response and fuzzy parameters minimizes the fuzziness of the model. The response observations of this model were the possibility of having the predefined property ($\mu_i = \text{Poss}(Y_i = 1) \epsilon [0,1]$). In the second paper they studied a fuzzy logistic least squares model with crisp input, fuzzy response and fuzzy parameters.

Takemura proposed a logistic regression analysis for fuzzy data in which input data, output data and parameters were all represented by L–R fuzzy numbers. He presented his fuzzy logistic model and used both possibilistic and fuzzy least squares approaches to estimate the parameters. Yang and Chen estimated the parameters of a logistic regression mixture using an algorithm called a fuzzy classification maximum likelihood (FCML) and then compared the mean squared error (MSE) of FCML and EM algorithms.

METHODOLOGY

Consider the situation in which the response variable is a fuzzy observation on the status of each case relative to binary response categories i.e. it takes two labels: approximately 1 or approximately 0 instead of 1 or 0. Due to the vague status of cases relative to response categories, the binary response observations are not precise.

So we cannot calculate the exact probability of success and the Bernoulli distribution is not helpful. A solution is to consider the possibility of success instead of the probability. Here, the possibility of success can be considered as a linguistic term such as very low, low, medium, high or very high (each of which is represented by a fuzzy number). Suppose the possibility of i'th sample case is $\tilde{p}_i$. Then we propose the following model:

$$\tilde{y}_i = \log\frac{\tilde{p}_i}{1-\tilde{p}_i} = \tilde{A}_0 + \tilde{A}_1 x_{i1} + \cdots + \tilde{A}_k x_{ik} + \tilde{\varepsilon}_i \ , i = 1, \dots, n \ ,$$

where x_{ij}'s are crisp inputs (whithout loss of generality, $x_{ij} > 0$), $\tilde{A}_j, j = 0,1,\dots,k$ are the fuzzy coefficients and $\tilde{\varepsilon}_i$ is the error without the distribution assumption. $\tilde{y}_i$ as the fuzzy output observation in fact is the logarithm transformation of possibilistic odds ($\tilde{y}_i = \log\frac{\tilde{p}_i}{1-\tilde{p}_i}$) in which $\tilde{p}_i$ is the possibility of success.

The membership function of $\tilde{y}_i$ is computed from the membership function of $\tilde{p}_i$ using extension principle.

Now we rewrite the model based on the α-cuts of fuzzy numbers:

$$y_{i\alpha} = A_{0\alpha} + A_{1\alpha} x_{i1} + \cdots + A_{k\alpha} x_{ik} + \varepsilon_{i\alpha} \ , i = 1, \dots, n \ , \alpha \epsilon [0,1]$$

For using the fuzzy least squares method in this model we need to minimize the sum of square errors between the α-cuts of observed outputs ($y_{i\alpha}$) and the α-cuts of estimated outputs ($\hat{y}_{i\alpha}$) i.e. $SSE_\alpha = \sum_{i=1}^{n} d^2(y_{i\alpha}, \hat{y}_{i\alpha})$, where d is the distance defined in (1).

To estimate α-cuts of parameters the partial derivatives are set to zero.

$$\frac{\partial SSE_\alpha}{\partial A_{j\alpha}^U} = 0 \ , \ \frac{\partial SSE_\alpha}{\partial A_{j\alpha}^L} = 0.$$

This leads to the following equations:

$$A_{0\alpha}^L \sum_{i=1}^{n} x_{i0} x_{ij} + A_{1\alpha}^L \sum_{i=1}^{n} x_{i1} x_{ij} + \cdots + A_{k\alpha}^L \sum_{i=1}^{n} x_{ik} x_{ij} = \sum_{i=1}^{n} y_{i\alpha}^L x_{ij} \ , \qquad j = 0,1,\dots,k$$

and

$$A_{0\alpha}^U \sum_{i=1}^{n} x_{i0} x_{ij} + A_{1\alpha}^U \sum_{i=1}^{n} x_{i1} x_{ij} + \cdots + A_{k\alpha}^U \sum_{i=1}^{n} x_{ik} x_{ij} = \sum_{i=1}^{n} y_{i\alpha}^U x_{ij} \ , \qquad j = 0,1,\dots,k$$

where

$x_{i0} = 1 \ , i = 1,2,\dots,n$

$A_{j\alpha} = [A_{j\alpha}^L, A_{j\alpha}^U] \ , j = 0,1,\dots,k$

$y_{i\alpha} = [y_{i\alpha}^L, y_{i\alpha}^U] \ , i = 1,2,\dots,n.$

These equations can be represented in the matrix form

$$(\mathbf{X}^T\mathbf{X})\mathbf{A}^L = \mathbf{X}^T\mathbf{Y}^L \ , \ (\mathbf{X}^T\mathbf{X})\mathbf{A}^U = \mathbf{X}^T\mathbf{Y}^U$$

in which

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}$$

$\mathbf{A}^L = (A^L_{0\alpha}, A^L_{1\alpha}, \dots, A^L_{k\alpha})^T$, $\mathbf{A}^U = (A^U_{0\alpha}, A^U_{1\alpha}, \dots, A^U_{k\alpha})^T$
and
$\mathbf{Y}^L = (y^L_{1\alpha}, y^L_{2\alpha}, \dots, y^L_{n\alpha})^T$, $\mathbf{Y}^U = (y^U_{1\alpha}, y^U_{2\alpha}, \dots, y^U_{n\alpha})^T$.
If matrix the $\mathbf{X}^T\mathbf{X}$ is positive definite then $(\mathbf{X}^T\mathbf{X})^{-1}$ is computable and the minimization problem has a unique solution as follows:
$\mathbf{A}^L = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}^L$, $\mathbf{A}^U = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}^U$.
We write $\widehat{A}_{j\alpha} = \left[\min\{\widehat{A}^L_{j\alpha}, \widehat{A}^U_{j\alpha}\}, \max\{\widehat{A}^L_{j\alpha}, \widehat{A}^U_{j\alpha}\}\right], j = 0,1, \dots, k.$
Having the α-cuts of parameters estimators, one can use the resolution identity introduced by Zadeh (1975) and obtain the fuzzy estimators of parameters as:

$$\widehat{\widehat{A}}_j = \bigcup_{\alpha\epsilon[0,1]} \alpha\widehat{A}_{j\alpha}$$

where $\widehat{A}_{j\alpha}$ is the α-cut estimator of the unknown parameter.
Goodness-of-fit
The goodness-of-fit between the observed α-cutvalues and the estimated α-cut values obtained by the model are evaluated using a capability index.
Definition 4**:** Suppose $A, B \in F(X)$. Then, the capability index between α-cuts of A and B is defined by Taheri and Kelkinnama (2008).

$$I_{UI\alpha}(A, B) = \frac{\mathrm{Card}(A_\alpha \cap B_\alpha)}{\mathrm{Card}(A_\alpha \cup B_\alpha)}$$

where $\mathrm{Card}(A_\alpha)$ is the length of the interval A_α.
For fuzzy logistic regression, the mean of the capability index is used as a measure of the goodness-of-fit of the model.
$\mathrm{MCI}_\alpha = \frac{1}{n}\sum_{i=1}^n I_{UI\alpha}(y_{i\alpha}, \hat{y}_{i\alpha})$.
We always have $0 \le \mathrm{MCI}_\alpha \le 1$, and the larger the MCI_α, the better the goodness-of-fit.

Bootstrap Fuzzy Logistic Regression Analysis

In this section we give a brief review of the bootstrap technique in regression analysis. Bootstrapping is a general approach to statistical inference based on building a sampling distribution for a statistic by resampling from the data at hand. The term "bootstrapping", first introduced by Efron (1979) is an allusion to the expression "pulling oneself up by one's bootstraps" (in this case, using the sample data as a population from which repeated samples are drawn). The key bootstrap analogy is therefore as follows: The population is to the sample as the sample is to the bootstrap samples. There are two general ways to bootstrap a regression model. Assume that we want to fit a regression model with response variable$y$ and a predictor x. The predictor can be supposed as random, potentially changing from sample to sample, or as fixed. In the first case (random-x) it is better to resample from observations. For this purpose assume that we have a sample of n observations $z_i = (x_i, y_i)$, $i = 1,2, \dots, n$, then we simply select bootstrap samples of the z_i's. In the second case (fixed-x) the method proceeds by resampling errors. Attaching resampling errors to each fitted value from the model $(\hat{y}_i)$ produces a bootstrap sample.
In this paper we use the second method which proceeds according to the following algorithm:
Step1: Fit the fuzzy LS model and obtain the estimated response α-cuts as $\hat{y}_{i\alpha} = \widehat{A}_{0\alpha} + \widehat{A}_{1\alpha}x_{i1} + \cdots + \widehat{A}_{k\alpha}x_{ik}$ and calculate the residuals as: $\hat{\varepsilon}_{i\alpha} = y_{i\alpha} \ominus_g \hat{y}_{i\alpha}$.
Step2: Denote the centered residuals by $e_{i\alpha} = \hat{\varepsilon}_{i\alpha} \ominus_g \bar{\hat{\varepsilon}}_{i\alpha}$ where $\bar{\hat{\varepsilon}}_{i\alpha}$ is the mean of $\hat{\varepsilon}_{i\alpha}$'s $(i = 1, \dots, n)$.
Step3: Let $\hat{F}_n$ be the empirical distribution of residuals, centered at the mean, so that $\hat{F}_n$ puts mass $\frac{1}{n}$ at each $e_{i\alpha}$, then generate a sample of $e_{i\alpha}$ from $\hat{F}_n$. (Informally, draw an n-sized bootstrap random sample with replacement from the $e_{i\alpha}$). Write these new centered residuals as $e^b_{i\alpha}$, $i = 1, \dots, n$then the bootstrap sample is generated by: $y^b_{i\alpha} = \hat{y}_{i\alpha} + e^b_{i\alpha}$.
Step4: Having this bootstrap sample , fit the fuzzy least squares model and obtain the estimates as $\widehat{A}^b_{j\alpha}$, $j = 0,1, \dots, k$.
Step5: Repeat step 3 and step 4 for a large enough number B.

Resampling residuals in this way and randomly reattaching them to fitted values ensures that errors are identically distributed.

Bootstrapping draws an analogy between the fitted value $\hat{y}_\alpha$ in the sample and y_α in the population, and between the residual e_α in the sample and the error ε_α in the population. According to the weak law of large numbers, the empirical distribution function $\hat{F}_n$ converges in probability to the true distribution function. Note that we define the bootstrap observation $y_{i\alpha}^b$, by treating $\widehat{A}_{j\alpha}$ as the true parameter and $e_{i\alpha}^b$ as the population of errors.

The Bootstrap-based Hypotheses Test

It is also possible to use the bootstrap method to construct an empirical sampling distribution for a test statistic. Classical hypothesis testing methods are usually based on the statistics whose distributions depend on the distribution of errors. However the bootstrap techniques use the empirical distribution of the test statistic and does not need any distribution assumption. To construct a bootstrap test of the hypothesis $H_0: A_{j\alpha} = A_{j\alpha}^*$ vs $H_1: A_{j\alpha} \neq A_{j\alpha}^*$, we use the test statistic proposed by Lee et al. as follows:

$$T_{j\alpha} = \sqrt{d^2\left(\frac{\widehat{A}_{j\alpha} \ominus_g A_{j\alpha}^*}{s_{\widehat{A}_{j\alpha}}}, \{0\}\right)} \qquad (2)$$

where d is the distance defined in (1),

$$s_{\widehat{A}_{j\alpha}}^2 = g[j+1] \frac{\sum_{i=1}^{n} d^2(y_{i\alpha}, \hat{y}_{i\alpha})}{n-2}$$

And $g[i]$ is the jth diagonal entry of the matrix $(\mathbf{X}^T\mathbf{X})^{-1}$.

Now, using (2), we can calculate the bootstrap test statistic $T_{j\alpha}^b$ for each bootstrap sample $(b = 1, \ldots, B)$. Using the empirical distribution of $T_{j\alpha}^b$, we can compute the p_value of the test as the proportion of values $\{T_{j\alpha}^1, \ldots, T_{j\alpha}^B\}$ greater than or equal to the main test statistic $T_{j\alpha}$.

$$\text{p\_value}_\alpha = \frac{\#\{b: T_{j\alpha}^b \geq T_{j\alpha}\}}{B}$$

Now suppose the level of significance of the hypothesis test is denoted by λ . We reject the null hypothesis H_0 at significance level λ if $\text{p\_value}_\alpha < \lambda$ with $(1 - \text{p\_value}_\alpha)$ degree of rejection.

Now we compute $100(1-\lambda)$ % confidence region for parameters, using the empirical sampling distribution of $T_{j\alpha}^b$.

Based on the ordered values of $T_{j\alpha}^b, j = 0,1, \ldots, k$ (or quantiles of the bootstrap empirical distributions), we obtain the critical points $t_{j\lambda}$, $j = 0, \ldots, k$ such that $P(T_{j\alpha} < t_{j\lambda}) = 1 - \lambda$. Therefore the estimated confidence regions are computed as:

$$\left(A_{j\alpha}^C - \widehat{A}_{j\alpha}^C\right)^2 + \left(A_{j\alpha}^W - \widehat{A}_{j\alpha}^W\right)^2 < t_{j\alpha}^2 s_{\widehat{A}_{j\alpha}}^2 , \; j = 0, \ldots, k \qquad (3)$$

where $A_{j\alpha}^C$ and $A_{j\alpha}^W$ are the midpoint and width of $A_{j\alpha}$, $j = 0, \ldots, k$ respectively.

Simulation Study and Numerical Example

In this section we present a simulation study to illustrate the effectiveness of our model. The simulation was carried out using Monte Carlo techniques.

Example 1**.** In this example, we simulated a sample of "x_i-values" from the uniform distribution on (0, 10). The "observed" response was chosen as:

$$\tilde{y}_i = \log\frac{\tilde{p}_i}{1-\tilde{p}_i} = \widetilde{A}_0 + \widetilde{A}_1 x_i + \varepsilon_i,$$

where $\varepsilon_i \overset{iid}{\sim} N(0,1)$, $\widetilde{A}_0$ varies over the triangular fuzzy numbers with centers 3, 5, 8 and respective spreads 1, 2, 3 and $\widetilde{A}_1$ varies over the triangular fuzzy numbers with centers 0.5, 0.8, 1.4, 2 and respective spreads 0.2, 0.3, 0.4, 0.5.

We investigated the finite sample performance of the estimators by varying the sample size within each simulation experiment. Specific sample sizes under consideration in this simulation are: 30, 50 and 100.

Applying the fuzzy least squares method described in section 3.2 we estimated the α-cuts of the parameters $\widetilde{A}_0$ and $\widetilde{A}_1$ for each sample $(\alpha = 0)$. Here we considered 1000 replications of each experiment and then computed the mean estimates, bias and MSE across the 1000 replicate simulations (Table 1). The formula for bias was as follows:

$$\text{Bias}(\widetilde{A}_{j\alpha}) = \frac{1}{1000} \sum_{i=1}^{1000} \left|d(\widetilde{A}_{j\alpha}, A_{j\alpha})\right|,$$

$$\mathrm{MSE}(\tilde{A}_{j\alpha}) = \frac{1}{1000}\sum_{i=1}^{1000} d^2(\tilde{A}_{j\alpha}, A_{j\alpha}) \quad , j = 0,1.$$

As seen in Table 1, when the sample size increases, the value of MSE and Bias decreases and the α-cut estimates of parameters tend to their real values.

Example 2. In this example we use a numerical example to illustrate the fuzzy logistic model that discussed in previous sections. The data set for this application was taken from Pourahamad et al .

Table 1.Mean Estimates, Bias and MSE of Fitted Model on Simulated Samples.

			$\hat{A}_{0\alpha}$			$\hat{A}_{1\alpha}$		
$A_{0\alpha}$	$A_{1\alpha}$	n	Mean	bias	MSE	Mean	bias	MSE
(2,4)	(0.3,0.7)	30	(1.9862,3.9853)	0.3332	0.1431	(0.3028,0.7029)	0.0580	0.0043
		50	(1.9961,4.0186)	0.2538	0.0819	(0.2997,0.6966)	0.0441	0.0025
		100	(1.9965,4.0008)	0.1817	0.0425	(0.3006,0.6993)	0.0315	0.0013
	(0.5,1.1)	30	(1.9829,4.0096)	0.3385	0.1456	(0.5022,1.0992)	0.0588	0.0044
		50	(2.0081,4.0171)	0.2543	0.0832	(0.5007,1.0981)	0.0443	0.0025
		100	(1.9994,3.9962)	0.1758	0.0398	(0.4994,1.1003)	0.0307	0.0012
	(1,1.8)	30	(2.0096,4.0085)	0.3322	0.1398	(0.9994,1.8005)	0.0591	0.0044
		50	(2.0091,3.9954)	0.2622	0.0885	(0.9994,1.8004)	0.0444	0.0025
		100	(1.9926,4.0027)	0.1808	0.0412	(1.0009,1.7989)	0.0305	0.0012
	(1.5,2.5)	30	(2.0181,3.9914)	0.3238	0.1367	(1.4972,2.5007)	0.0566	0.0041
		50	(2.0127,4.0031)	0.2499	0.0798	(1.4981,2.5001)	0.0432	0.0024
		100	(2.0051,3.9981)	0.1789	0.0406	(1.4987,2.4996)	0.0317	0.0013
(3,7)	(0.3,0.7)	30	(2.9942,7.0026)	0.3290	0.1392	(0.3021,0.6989)	0.0572	0.0041
		50	(3.0034,6.9972)	0.2529	0.0801	(0.3005,0.6999)	0.0443	0.0025
		100	(3.0025,6.9915)	0.1807	0.0415	(0.2991,0.7007)	0.0311	0.0012
	(0.5,1.1)	30	(2.9898,6.9715)	0.3390	0.1478	(0.5010,1.1047)	0.0584	0.0044
		50	(2.9959,7.0002)	0.2545	0.0830	(0.5003,1.0984)	0.0444	0.0025
		100	(2.9938,6.9977)	0.1830	0.0420	(0.5020,1.1002)	0.0312	0.0012
	(1,1.8)	30	(2.9848,7.0069)	0.3311	0.1401	(1.0024,1.7996)	0.0578	0.0042
		50	(3.0026,7.0134)	0.2562	0.0837	(1.0012,1.7993)	0.0448	0.0025
		100	(2.9976,6.9972)	0.1774	0.0406	(0.9999,1.8004)	0.0308	0.0012
	(1.5,2.5)	30	(3.0290,7.0226)	0.3138	0.1305	(1.4966,2.4966)	0.0555	0.0039
		50	(3.0040,6.9952)	0.2613	0.0878	(1.5006,2.4997)	0.0447	0.0026
		100	(2.9973,6.9994)	0.1804	0.0408	(1.4999,2.4993)	0.0304	0.0011
(5,11)	(0.3,0.7)	30	(4.9850,11.0141)	0.3339	0.1433	(0.3038,0.6979)	0.0567	0.0041
		50	(5.0110,11.0199)	0.2515	0.0818	(0.2986,0.6980)	0.0441	0.0025
		100	(4.9987,10.9981)	0.1787	0.0405	(0.2994,0.6997)	0.0308	0.0012
	(0.5,1.1)	30	(5.0021,11.0163)	0.3200	0.1321	(0.5005,1.0963)	0.0562	0.0040
		50	(5.0179,11.0058)	0.2562	0.0828	(0.4965,1.0989)	0.0443	0.0024
		100	(5.0092,10.9964)	0.1768	0.0407	(0.4988,1.1013)	0.0305	0.0012
	(1,1.8)	30	(4.9942,10.9971)	0.3291	0.1398	(1.0009,1.8001)	0.0580	0.0043
		50	(5.0052, 10.9876)	0.2488	0.0794	(0.9995, 1.8024)	0.0435	0.0023
		100	(4.9911,10.9974)	0.1765	0.0401	(0.9999,1.7999)	0.0299	0.0011
	(1.5,2.5)	30	(4.9947,11.0113)	0.3327	0.1424	(1.4992,2.4974)	0.0579	0.0043
		50	(5.0147,11.0077)	0.2490	0.0804	(1.4996,2.4986)	0.0441	0.0025
		100	(4.9954,10.9993)	0.1768	0.0405	(1.5006,2.5006)	0.0307	0.0012

Lupus is a chronic (long-lasting) autoimmune disease where the immune system, for unknown reasons, createsantibodies(proteins produced by white blood cells) which, instead of protecting the body from bacteria and viruses, attack normal body tissues. Systematic Lupus Erythematous (SLE) is a lupus which attacks multiple systems in the body including the skin, blood, lungs, heart, brain and nervous system. In our point of view, the distinction borderline between patients and healthy people should not be considered crisp for SLE. Also, the degree of disease depends on the patient group. Therefore, to model the relationship between possibility odds of this disease and a set of explanatory variables, our proposed model is a good choice. Our sample contains 15 females aged 18-40 who are suspected to have SLE. An expert is asked to assign a possibility of disease to each case using a term like very low, low, medium, high, very high. Family history (X_1), sun exposure (X_2), and some diagnostic blood tests such as Anti Nuclear Antibody (ANA) test (X_3), Anti DNA test (X_4) and Erythrocyte Sedimentation Rate (ESR) test (X_5) are used as the significant risk factors (Table 2).

Table 2.The Sample Data

Obs.	X_1	X_2	X_3	X_4	X_5	Possibility of SLE
1	1	1	112	105	1	high
2	0	1	80	23	0	medium
3	0	1	115	15	0	high
4	0	1	105	107	1	high
5	0	0	89	150	1	medium
6	1	1	160	10	1	Very high
7	0	1	100	23	0	medium
8	0	0	100	85	1	high
9	0	1	48	83	0	low
10	1	0	15	19	1	very low
11	0	0	50	91	0	low
12	0	1	59	200	1	medium
13	0	1	83	20	1	low
14	0	0	15	200	0	low
15	1	0	85	15	1	medium

For this situation Pourahmad et al. assume the observations as triangular fuzzy numbers that cover the range of (0,1), see Figure1. To define the relationship between possibility odds of SLE disease and the mentioned risk factors, we use the following model:

$$\tilde{y}_{i=}\log\frac{\tilde{p}_i}{1-\tilde{p}_i}=\tilde{A}_0+\tilde{A}_1x_{i1}+\cdots+\tilde{A}_5x_{i5}+\tilde{\varepsilon}_i\ \ ,i=1,2,\ldots,15$$

The lower and upper bounds of the α-cuts of $\widehat{\tilde{A}}_j$, $j=0,1,\ldots,5$ for $\alpha=0,0.1,0.2,\ldots,0.9,1$ are provided in Table 3. $\alpha=1.0$ shows the regression coefficient that is most likely, and the $\alpha=0.0$ shows the range in which the regression coefficient could appear. For example, the most likely values of $\widehat{\tilde{A}}_0$, $\widehat{\tilde{A}}_1$ and $\widehat{\tilde{A}}_2$ are -4.345, 0.542 and -0.139, respectively and it is impossible for them to take outside the ranges of [-4.820, -2.969], [0.116,0.759] and [-0.167,-0.112], respectively. For case 1, as an

Table 3. Estimated $\boldsymbol{\alpha}$-cuts of $\tilde{\mathbf{A}}_0$ and $\tilde{\mathbf{A}}_1$ for Different Values of $\boldsymbol{\alpha}$

α	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\hat{A}_0^L$	-4.820	-4.678	-4.557	-4.456	-4.369	-4.299	-4.247	-4.213	-4.204	-4.234	-4.345
$\hat{A}_0^U$	-2.969	-3.018	-3.080	-3.154	-3.242	-3.343	-3.462	-3.604	-3.778	-4.005	-4.345
$\hat{A}_1^L$	0.116	0.119	0.126	0.137	0.154	0.176	0.206	0.246	0.303	0.388	0.542
$\hat{A}_1^U$	0.759	0.753	0.746	0.738	0.728	0.716	0.701	0.682	0.655	0.614	0.542
$\hat{A}_2^L$	-0.167	-0.162	-0.158	-0.155	-0.152	-0.149	-0.147	-0.145	-0.143	-0.141	-0.139
$\hat{A}_2^U$	-0.112	-0.110	-0.110	-0.109	-0.110	-0.111	-0.113	-0.116	-0.121	-0.127	-0.139
$\hat{A}_3^L$	0.0419	0.0415	0.0412	0.0412	0.0413	0.0416	0.0421	0.0429	0.0440	0.0457	0.0487
$\hat{A}_3^U$	0.0441	0.0436	0.0433	0.0432	0.0432	0.0434	0.0437	0.0443	0.0451	0.0464	0.0487
$\hat{A}_4^L$	0.0078	0.0077	0.0078	0.0078	0.0079	0.0080	0.0082	0.0085	0.0089	0.0094	0.0104
$\hat{A}_4^U$	0.0096	0.0095	0.0095	0.0095	0.0095	0.0096	0.0096	0.0097	0.0099	0.0101	0.0104
$\hat{A}_5^L$	-0.528	-0.545	-0.561	-0.576	-0.592	-0.610	-0.629	-0.651	-0.678	-0.715	-0.775
$\hat{A}_5^U$	-0.447	-0.459	-0.473	-0.488	-0.505	-0.525	-0.550	-0.580	-0.621	-0.678	-0.775

example, with $\alpha = 0$ the α-cut of estimated output based on our proposed model is:

$$\hat{y}_{1\alpha} = [-4.820, -2.969] + [0.116, 0.759] \times 1 + [-0.167, -0.112] \times 1 + [0.0419, 0.0441] \times 112 + [0.0078, 0.0096] \times 105 + [-0.528, -0.447] \times 1$$

so that

$\hat{y}_{1\alpha} = \left(\log\frac{\tilde{p}_1}{1-\tilde{p}_1}\right)_\alpha = [0.1128, 3.1782]$.

Now by one-to-one property of $f(x) = \frac{\exp(x)}{1+\exp(x)}$, the estimated α-cut of possibility of SLE for this case is:

$\hat{p}_{1\alpha} = \left[\frac{\exp(0.1128)}{1+\exp(0.1128)}, \frac{\exp(3.1782)}{1+\exp(3.1782)}\right] = [0.5282, 0.9600]$, $\alpha = 0$.

The observed α-cut of possibility of SLE for this case for $\alpha = 0$ is $[0.6, 0.9]$.

Now consider a new case with the information such as $x_1 = 1$, $x_2 = 0$, $x_3 = 110$, $x_4 = 87$, $x_5 = 0$. The estimated α-cut (for $\alpha = 0$) of output for this new case using our proposed model is:

$$\hat{y}_{0\alpha} = [-4.820, -2.969] + [0.116, 0.759] \times 1 + [-0.167, -0.112] \times 0 + [0.0419, 0.0441] \times 110 + [0.0078, 0.0096] \times 87 + [-0.528, -0.447] \times 0 = [0.5836, 3.4762]$$

So the estimated α-cut of possibility of disease for this case is:

$\hat{p}_{0\alpha} = [0.6419, 0.9700]$, $\alpha = 0$.

To evaluate the model based on the proposed index (Definition 4), we calculate MCI_α for various $\alpha \epsilon [0,1]$ and compare our model with the model proposed by Pourahmad et al.. As seen in Table 4, the MCI_α index of our model is greater than other's for all α‘s except $\alpha = 0.4, 0.5$.

Also, the mean of this index for a sequence of 1000 α's in our model is 0.512 and in other model is 0.499. To check whether linear relationship between the independent variables and the dependent variable is significant, we consider two hypotheses as follows:

$H_0: A_{j\alpha} = \{0\}$ $v.s.$ $H_1: A_{j\alpha} \neq \{0\}$, $j = 0,1,\dots,5$.

For α-cuts, the null hypothesis states that $A_{j\alpha}$ is equal to $\{0\}$, and the alternative hypothesis states that $A_{j\alpha}$ is not equal to $\{0\}$. To perform the test, 1000 replicate data sets were created by bootstrap method using the residuals. If the p_value is smaller than significance level λ, then the null hypothesis isrejected. Figure 2 shows the behavior of the p_value for various α-cuts of each A_j , $j = 0,1,\dots,5$. As shown in Figure 2, for A_0, A_3 and A_4 we reject the null hypothesis and for A_2 and A_1 we accept the null hypothesis at the level of significance $\lambda = 0.05$ for all α. Also for A_5 when α-cuts are larger than 0.1, we

Table 4. MCI_α for different values of α

α	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Our model	0.667	0.647	0.623	0.593	0.558	0.515	0.463	0.391	0.296	0.146
Pourahmad et al. model	0.631	0.619	0.604	0.587	0.563	0.524	0.461	0.380	0.260	0.090

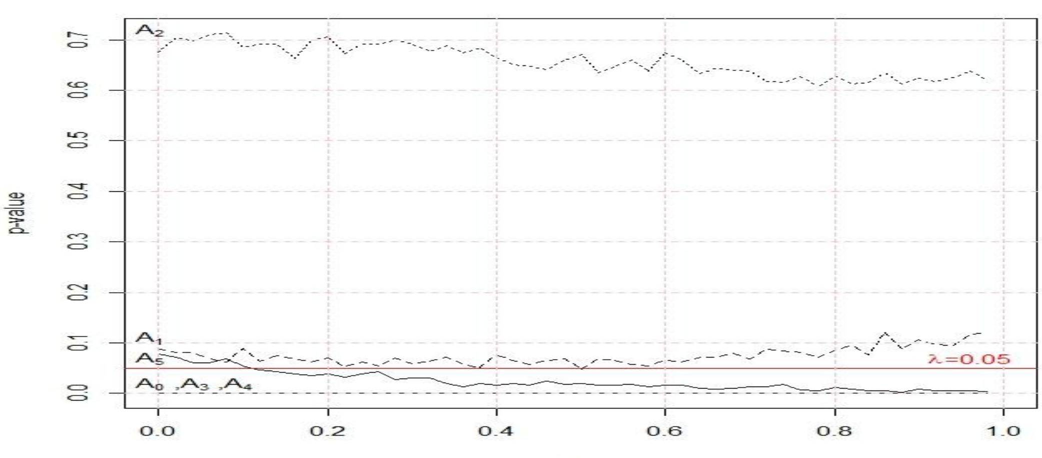

Figure 1. The Behavior of p-value in Testing $H_0: A_{j\alpha} = \{0\}$ v.s. $H_1: A_{j\alpha} \neq \{0\}$, $j = 0,1,\dots,5$

Reject the null hypothesis at the level of significance $\lambda = 0.05$ and when α-cuts are smaller than 0.1, we accept the null hypothesis. This example shows that the statistical significance of the coefficients changes depending on the vagueness of the data.

The 95% confidence region defined in (3) for A_0, A_1, A_2andA_5 is provided in Figure 3, where the x-axis denotes the center $(A_{j\alpha}^{C} = \frac{A_{j\alpha}^{U}+A_{j\alpha}^{L}}{2})$ of $A_{j\alpha} = [A_{j\alpha}^{L}, A_{j\alpha}^{U}]$, the y-axis represents the width $(A_{j\alpha}^{W} = \frac{A_{j\alpha}^{U}-A_{j\alpha}^{L}}{2})$ of $A_{j\alpha} = [A_{j\alpha}^{L}, A_{j\alpha}^{U}]$ and z-axis is α.

CONCLUSION

In this paper, we have presented an adaptive fuzzy logistic regression model based on the least squares method. The proposed model has an advantage compared to other adaptive fuzzy logistic regression models; this model estimates parameters for each α-cut of fuzzy observations of various kinds. On the other hand there is no need that observations be a special type of fuzzy numbers (such as L-R or triangular fuzzy numbers). It is enough to have α-cuts of fuzzy numbers and the membership function is obtained using the resolution identity.

The proposed model is recommended for crisp input and fuzzy binary output observations. This adaptive model uses the logarithmic transformation of possibility of success $(\tilde{p}_i)$ for each case.We consider $\tilde{p}_i$as a linguistic term (very low, low, medium, high and very high) by assigning a triangular fuzzy number to each output in such a way that the union of their supports covers the whole range of (0,1) interval.

We also discuss statistical inference in the presence of fuzzy data using bootstrap techniques. For each α-cut, we tested hypotheses for the logistic regression model based on the fuzzy least squares estimator. A numerical example illustrates the hypothesis test and the confidence regions.

For future work the proposed approach can be generalized to prediction problems involving other types of intrinsic linear functions (such as exponential, reciprocal and growth functions) in the fuzzy environment with suitable linear transformations.

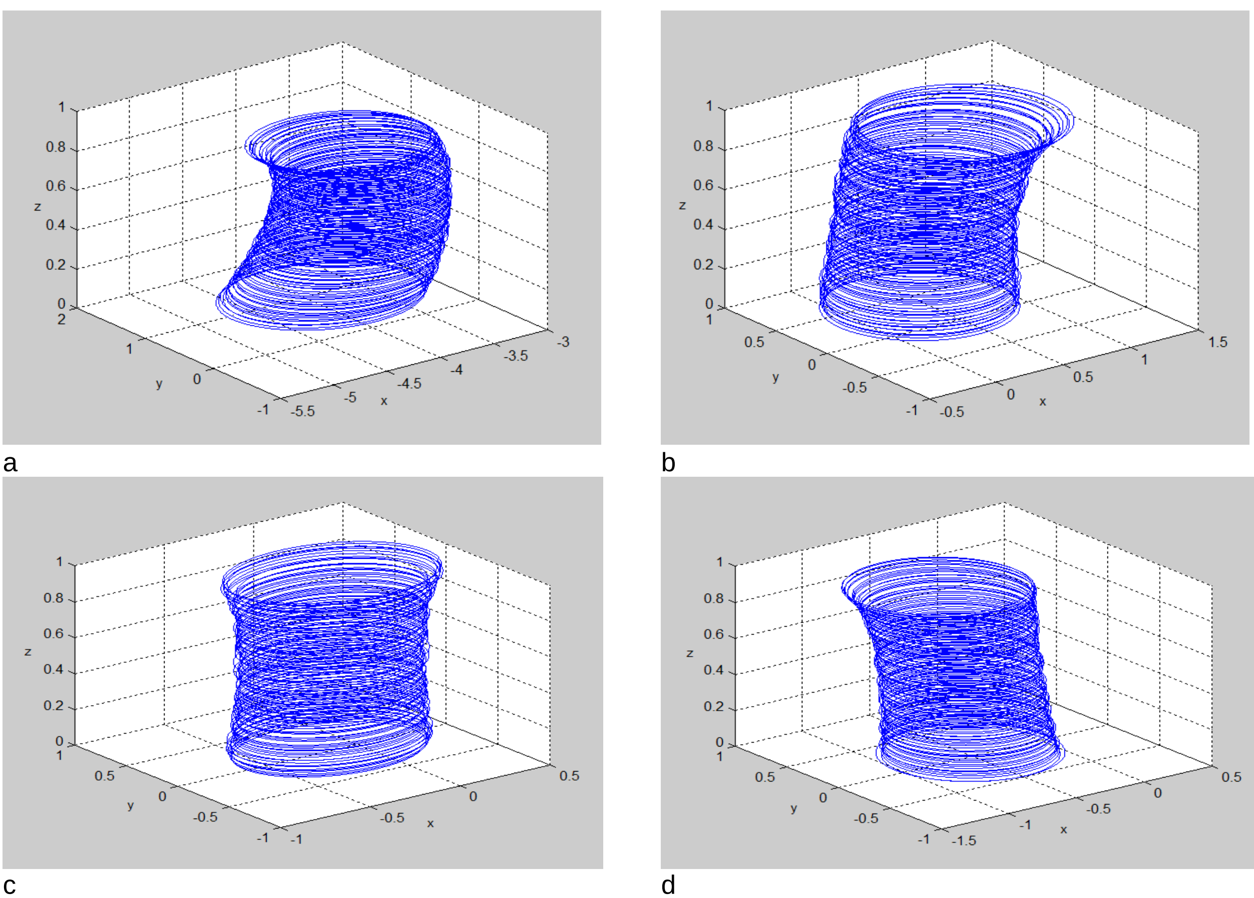

Figure 2. 95% Confidence Regions of a: $\mathbf{A_0}$ b: $\mathbf{A_1}$ c: $\mathbf{A_2}$d: $\mathbf{A_5}$

REFERENCES

Celmins A.: Least squares model fitting to fuzzy vector data, Fuzzy Sets and Systems 22 (1987) 245-269.

Celmins A.: Multidimensional least-squares model fitting offuzzy models, Mathematical Modeling 9 (1987) 669-690.

Chang Y.O.,Ayyub B.M.: Fuzzy regression methods: a comparative assessment, Fuzzy Sets and Systems 119 (2001) 187-203.

Diamond P.: Fuzzy least square, Information Sciences 46 (1988) 141-157.

Dom R.M., Zain R., Abdul Kareem S., Abidin B.: An adaptive fuzzy regression model for the prediction of dichotomous response variables, fifth International Conference on Computational Science and Applications, Malaysia (2007) pp.14-19.

Dubois D., Prade H.: Possibility Theory, Plenum Press, New York, (1988).

Efron B.: Bootstrap methods: another look at the Jackknife, Annals of Statistics 7 (1979) 1-26.

Efron B.: The jackknife, the bootstrap, and other resampling plans,vol 38. Society of Industrial and Applied Mathematics CBMS-NSF Monographs, Philadelphia (1982) ISBN 0898711797.

Lee W.J., Jung H.Y., Yoon J.H., Choi S.H.: The statistical inferences of fuzzy regression based on bootstrap techniques, Soft Comput 19 (2015) 883-890.

Nagar P., Srivastava S.: Adaptive fuzzy regression model for the prediction of dichotomous response variables using cancer data: a case study, Journal of Applied Mathematics, Statistics and Informatics 4 (2008) 183-191.

Pourahmad S., Ayatollahi S.M.T., Taheri S.M. , Agahi Z.H.: Fuzzy logistic regression based on the least squares approach with application in clinical studies, Computers and Mathematics with Applications 62 (2011) 3353-3365.

Pourahmad S., Ayatollahi S.M.T., Taheri S.M.: Logistic regression, a new posssibilistic regression and its application in clinical vague status, Iranian Journal of Fuzzy Systems 8 (2011) 1-17.

Stefanini L.: A generalization of Hukuhara difference for intervaland fuzzy arithmetic, Fuzzy Sets and Systems 161 (2010) 1564-1576.

Taheri S.M., Kelkinnama M.: Fuzzy least absolutes regression, The Proceedings of 4th International IEEE Conference on Intelligent Systems, Varna, Bulgaria 11 (2008) pp. 55-58.

Takemura K.: Fuzzy logistic regression analysis for fuzzy input and output data, The Proceedings of the Second International Symposium on Soft Computing and Intelligent System, Japan (2004) pp. 1-6.

Tanaka H., Uejima S., Asai K.: Regression analysis with fuzzy model, IEEE Transactions on Systems, Man and Cybernetics 12 (1982) 903-907.

Wu C.F.J.: Jackknife, bootstrap and other resampling methods in regression analysis, Annals of Statistics 14(4) (1986) 1261-1295.

Yang M.S., Chen H.M.: Logistic regression analysis, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 12 (2004) 761-780.

Zadeh L.A.: Fuzzy Sets, Information and control 8 (1965) 338-359.

Zadeh L.A.: The concept of linguistic variable and its application to approximate reasoning I, InfSci 8 (1975a) 199-249.

Zadeh L.A.: The concept of linguistic variable and its application to approximate reasoning II, InfSci 8 (1975b) 301-357.

Zadeh L.A.: The concept of linguistic variable and its application to approximate reasoning III, *InfSci 9* (1975c) 43-80.