Journal of Engineering and Applied Sciences 11 (9): 2105-2112, 2016
ISSN: 1816-949X

A Bootstrap-Based Method of Statistical Inference in Fuzzy Logistic Regression

[1]Farzaneh Moradi, [2]Alireza Arabpour and [1]Ali Shadrokh
[1]Department of Statistics, Payame Noor University, Tehran, Iran
[2]Department of Statistics, Faculty of Mathematics and Computer,
Shahid Bahonar University of Kerman, Kerman, Iran

Abstract: Statistical logistic regression is used for modeling a binary response variable based on a set of explanatory variables. In practice, the state of the response variable may be described in linguistic terms rather than in exact ones. So, it is not possible to categorize the samples in one of two response categories and no usual probability distribution can be considered for such binary response variables. Therefore, statistical logistic regression is not appropriate for modeling. In this study, researchers propose an adaptive fuzzy least squares model based on possibility of success that is defined by some linguistic terms. Also for each α-cut, using bootstrap technique, researchers discuss the problem of statistical inference. To evaluate the goodness of fit, a criterion named the capability index is calculated. At the end, because of widespread applications of logistic regression in clinical studies and also, the abundance of vague observations in clinical diagnosis, the suspected cases to Systematic Lupus Erythematosus (SLE) disease is modeled via an explanatory variable to detect the application of the model. The results showed that the proposed model could be a rational substituted model of an ordinary one in modeling the clinical vague status.

Key words: Fuzzy regression, logistic regression, least squares method, adaptive model, bootstrap, possibilistic odds, capability index, hypothesis testing

INTRODUCTION

Fuzzy linear regression models are used to obtain and appropriate linear relation between a dependent variable and several independent variables in a fuzzy environment. There are two categories of fuzzy regression analysis. The first is the possibilistic method which minimizes the total vagueness of the estimated values for the dependent variables. This analysis was first proposed by Zadeh *et al.* (1975a, b).

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 (1987a, b) and Diamond *et al.* (1988).

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, non-precise or vague observations are occurred frequently in practical situations, specially in medical studies. For example, due to lack of suitable instruments or well-defined criteria we may have suspicion in determining the state of the response variable (0 or 1) and therefore cannot categorize the individual samples in one of two response categories. Lupus is the example in this field which there is no biological examination and the disease is diagnosed by some defined and wholly accepted criteria. To distinguish patients in this disease, cases which have some of those defined criteria (not all of them) have a vague status. In addition, in some practical situations, it is more flexible and common to express the amount of response variable by linguistic terms such as “very low, low, average, high, very high” instead of crisp numbers. Therefore, in these situations due to the vague status of cases relative to response categories, a probability distribution cannot be considered for the response variable. If the size of data is large enough and proper basic assumptions are satisfied, ordinary logistic regression would be effective to analyze given data set. But, if data set includes ambiguous data which cannot be expressed by exact real number, fuzzy logistic regression could be an alternative choice (Namdari *et al.*, 2015).

In this study, 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

Corresponding Author: Farzaneh Moradi, Department of Statistics, Payame Noor University, Tehran, Iran

statistical inference based on building a sampling distribution for a statistic by resampling from the data at hand. Researchers shall also discuss estimation and hypothesis testing for the parameter of a fuzzy logistic regression model using bootstrap technique. There have been few studies on the bootstrap statistical inferences of fuzzy regression (Akbari *et al*., 2012; Lee *et al*., 2015).

Preliminaries: Some basic definitions are explained in this section. A fuzzy subset of a universe set X is specified by a membership function μ_A: X→[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→Y) and A∈F(X) then the extension principle allows us to define a fuzzy set B = f(A) in Y with following membership function:

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

Definition 1: A fuzzy subset of the set of real numbers R with membership function μ_A: R→[0, 1] is called a fuzzy number if:

- A is normal, i.e., there exists an element z_0 such that $\mu_A(z_0) = 1$
- A is fuzzy convex, i.e., $\forall_{z1}$, $z_2 \in R$ $\mu_A(\lambda z_1 + (1-\lambda) z_2) \geq \mu_A(z_1) \wedge \mu_A(z_2)$, $\forall \lambda \in [0, 1]$
- μ_A is upper semi-continuous
- Supp(A) = {z∈R: $\mu_A(z)>0$} is bounded

Definition 2: A fuzzy number can be represented as a family of sets called-cuts, A_a, defined as:

$$A_\alpha = \{z \in R: \mu_A(z) \geq \alpha\} \text{ for } 0<\alpha \leq 1$$

and:

$$A_\alpha = \{z\in R: \mu_A(z)>0\}$$

for α = 0 based on the resolution identity we get $A = U_{\alpha\in[0,1]} \alpha A_\alpha$. From the definition of fuzzy number it is easily seen that every -cut of a fuzzy number is a closed interval $A_\alpha = [A^l_\alpha, A^U_\alpha]$ where:

$$A^L_\alpha = \inf\{z\in R: \mu_A(z) \geq \alpha\},$$
$$A^u_\alpha = \sup\{z\in R: \mu_A(z) \geq \alpha\}$$

Definition 3: The arithmetic operations for two fuzzy numbers A and B with a-cuts $A = [A^L_\alpha, A^u_\alpha]$ and $B_\alpha = [B^l_\alpha, B^u_\alpha$ are defined as follow:

$$A_\alpha + B_\alpha = \left[A^L_\alpha + B^L_\alpha, A^U_\alpha + B^U_\alpha\right]$$

For given k∈R:

$$k.A_\alpha = \begin{cases} [kA^L_\alpha, kA^U_\alpha] \text{ if } k \geq 0 \\ [kA^U_\alpha, kA^L_\alpha] \text{ if } k<0 \end{cases}$$

$$k+A_\alpha = [k+A^L_\alpha, k + A^U_\alpha]$$

For the subtraction, we use the general Hukuhara difference:

$$\left[A^L_\alpha, A^U_\alpha\right]\Theta_g\left[B^L_\alpha, B^U_\alpha\right] = \left[C^-_\alpha, C^+_\alpha\right]$$

where, $C^-_a = \min\{A^L_\alpha - B^L_\alpha, A^u_\alpha - B^U_\alpha\}$ and $C^+_\alpha = \max\{A^l_\alpha - B^L_\alpha, A^u_\alpha - B^U_\alpha\}$. We can also represent an a-cut, A_α, by its midpoint and width, i.e., $A_\alpha = (A^C_\alpha, A^w_\alpha)$ where, $A^c_\alpha = 1/2 (A^L_\alpha + A^u_\alpha)$ and $A^w_\alpha = ½ (A^u_\alpha - A^L_\alpha)$. In this study, we define the Euclidean distance between cuts of two fuzzy numbers A and B (Lee *et al*., 2015) as:

$$d(A_\alpha, B_\alpha) = \sqrt{\left(A^C_\alpha - B^C_\alpha\right)^2 + \left(A^W_\alpha - B^W_\alpha\right)^2}$$

MATERIALS AND METHODS

Statistical inferences for fuzzy logistic regression model 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 Bernoulii with success probabilit 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. So the logit transformation is defined as logit(p) = log p/1-p. This function spreads the probabilities over the real number line (R) (and hence is a suitable link function. Our logistic regression model may now be written as follows:

$$\log\frac{p_i}{1-p_i} = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \varepsilon_i, i = 1, 2, ..., n$$

In such models the parameters are estimated by maximum likelihood methods. 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 (Phibanchon *et al*., 2007; Nagar and Srivastava, 2008) 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*. (2011a, b) 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. In the second study, they studied a fuzzy logistic least squares model with crisp input, fuzzy response and fuzzy parameters.

Hauser *et al*. (2012) introduced a classification method combining logistic regression and fuzzy logic in the determination of sampling sites for feral fish. Recently, Namdari *et al*. (2015) presented a fuzzy logistic regression model using Least Absolute Deviations (LAD) method. Chen and Wei (2016) developed a least squares model in fuzzy logistic regression with crisp input and fuzzy output that the coefficients and outputs are LR fuzzy numbers.

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). We propose the following model:

$$\tilde{y}_i = \log p_i \frac{\tilde{p}_i}{1-\tilde{p}_i} = \tilde{A}_0 + \tilde{A}_1 x_i + \tilde{\varepsilon}_i, i = 1, 2, ..., n$$

In this study, we restrict the model in to the univariate case without loss of generality; x_i is crisp input (without loss of generality, $x_i>0$), A_0 and A_1 are the fuzzy coefficients and ε_i is the error without the distribution assumption. y_i as the fuzzy output observation in fact is the logarithm transformation of possibilistic odds (y_i = log $p_i/1\text{-}p_i$) in which p_i is the possibility of success. The membership function of is computed from the membership function of using extension principle. Now we rewrite the model based on the-cuts of fuzzy numbers:

$$y_{i\alpha} = A_{0\alpha} + A_{1\alpha} x_i + \varepsilon_{i\alpha}, i = 1, 2, ..., n, \quad \alpha \in [0, 1]$$

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_{ia}) and the α-cuts of estimated outputs (y_{ia}) i.e., SSE = $\Sigma^n_{i=1} d^2(y_{ia}, y_{ia})$ where d is the distance defined in 1. To estimate α-cuts of parameters the partial derivatives are set to zero. By solving the obtained equations the -cut estimators for the unknown parameters are:

$$\hat{A}^L_{1\alpha} = \frac{S^L_{xy}}{S_{xx}}, \hat{A}^L_{0\alpha} = \bar{y}^L_\alpha - \hat{A}^L_{1\alpha}\bar{X}$$

and:

$$\hat{A}^U_{1\alpha} = \frac{S_{xy}U}{Sxx}, \hat{A}^U_{0\alpha} = \bar{y}^U_\alpha - \hat{A}^U_{1\alpha}\bar{X}$$

where, $S_{xx} = \Sigma^n_{i=1}(x_i\text{-}x)^2$ and $S_{xy}{}^L = \Sigma^n_{i=1}(y^L{}_{ia}\text{-}y^L{}_\alpha)$ (xi.x) and S_{xy} U = $\Sigma^n_{i=1}(y_{i\alpha}{}^U\text{-}y^U{}_\alpha)(x_i\text{-}x)$. Researchers write $A_{j\alpha}$ = [min $\{A^L_{j\alpha}, A^u_{j\alpha}\}$, max$\{A^L_{j\alpha}, A^u_{j\alpha}\}$], j = 0, 1 .Having the α-cuts of parameters estimators, one can use the resolution identity introduced by Zadeh (1975a, b) and obtain the fuzzy estimators of parameters as:

$$\hat{\hat{A}}_j = \cup_{\alpha\in[0,1]} \alpha \hat{A}_{j\alpha}$$

where, $\hat{A}_{j\alpha}$ is the α-cut estimator of the unknown parameter.

Goodness-of-fit: The goodness-of-fit between the observed α-cut values and the estimated α-cut values obtained by the model are evaluated using a capability index.

Definition 4: Suppose A, B∈F(X). Then, the capability index between α-cuts of A and B is defined by:

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

where, Card (A_α) 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:

$$MCI_\alpha = \frac{1}{n}\sum_{i=1}^{n} I_{UI\alpha}(y_{i\alpha}, \hat{y}_{i\alpha})$$

We always have $0 \leq MCI\alpha \leq 1$ and the larger the, the better the goodness-of-fit.

Bootstrap fuzzy logistic regression analysis: In this study, researchers 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. In this study we use the following algorithm a.

Alogrithm a:

Step 1: Fit the fuzzy LS model and obtain the estimated response α-cuts as $\hat{y}_{i\alpha} = \hat{A}_{0\alpha} + \hat{A}_{1\alpha}x_i$ and calculate the residuals as $\hat{\varepsilon}_{i\alpha} = y_{i\alpha} \Theta_g \hat{y}_{i\alpha}$.

Step 2: Denote the centered residuals by $e_{i\alpha}$, $e_{i\alpha} = \bar{\hat{\varepsilon}}_{i\alpha} \Theta_g \bar{\hat{\varepsilon}}_{i\alpha}$ where $\bar{\hat{\varepsilon}}_{i\alpha}$ is the mean of $\hat{\varepsilon}_{i\alpha}$'s (i = 1, ..., n).

Step 3: Let $\hat{F}n$ be the empirical distribution of residuals, centered at the mean, so that $\hat{F}n$ puts mass 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, ..., n then the bootstrap sample is generated by:

$$y^b_{i\alpha} = \hat{y}_{i\alpha} + e^b_{i\alpha}$$

Step 4: Having this bootstrap sample, fit the fuzzy least squares model and obtain the estimates as $\hat{A}^b_{0\alpha}$ and $\hat{A}^b_{1\alpha}$.

Step 5: Repeat step 3 and step 4 for a large enough number B. 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 $y^b_{i\alpha}$ by treating $\hat{A}_{j\alpha}$ as the true parameter and $e^b_{i\alpha}$ as the population of errors (Wu, 1986).

The bootstrap-based hypothesis 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: $A_{1\alpha} = A^*_{1\alpha}$ vs H_1: $A_{1\alpha} \neq A^*_{1\alpha}$, we use the test statistic proposed by Lee *et al*. (2015) as follows:

$$T_{1\alpha} = \sqrt{d^2\left(\frac{\hat{A}_{1\alpha}\Theta_g A^*_{1\alpha}}{S_{\hat{A}1\alpha}}, \{0\}\right)} \quad (1)$$

where, d is the distance defined in Eq. 1:

$$S^3_{\hat{A}1\alpha} = \frac{\sum_{i=1}^{n} d^2(y_{i\alpha}, \hat{y}_{i\alpha})}{(n-2)\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Now, using Eq. 1, researchers can calculate the bootstrap test statistic $T^b_{1\alpha}$ for each bootstrap sample (b = 1, ..., B). Using the empirical distribution of $T^b_{1\alpha}$ we can compute the p-value of the test as the proportion of values $\{T^1_{1\alpha}, ..., T^B_{1\alpha}\}$ greater than or equal to the main test statistic $T_{1\alpha}$:

$$\text{p-value} = \frac{\#\{b: T^2_{1\alpha} \geq T_{1\alpha}\}}{B}$$

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 p<λ with (1-p-value) degree of rejection. Also we can construct a bootstrap hypothesis test for the intercept coefficient as:

$$H_0: A_{0\alpha} = A^*_{0\alpha} \text{ vs } H_1: A_{0\alpha} \neq A^*_{0\alpha}$$

Here the same process as the test of is used. So, the test statistic is:

$$T_{0\alpha} = \sqrt{d^2\left(\frac{\hat{A}_{0\alpha}\Theta_g A^*_{0\alpha}}{S_{\hat{A}0\alpha}}, \{0\}\right)} \quad (2)$$

Where:

$$S^2_{\hat{A}1\alpha} = \frac{\sum_{i=1}^{n} d^2(y_{i\alpha}, \hat{y}_{i\alpha})}{(n-2)} \left(\frac{1}{n} + \frac{x^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$$

and then the bootstrap test statistic $T^b_{0\alpha}$ is computed for each bootstrap sample (b = 1, ..., B) and p-value is calculated as:

$$\text{p-value} = \frac{\#\{b: T^b_{0\alpha} \geq T_{0\alpha}\}}{B}$$

Now we compute 100(1-λ)% confidence region for the intercept and slope, using the empirical sampling distribution of $T^b_{0\alpha}$ and $T^b_{1\alpha}$. Based on the ordered values of $T^b_{j\alpha}$, j = 0, 1 (or quantiles of the bootstrap empirical distributions), researchers e obtain the critical points $t_0\lambda$ and $t_{1\lambda}$ such that $P(T_{0\alpha}<t_{0\lambda}) = p(T_{1\alpha}<t_{1\lambda}) = 1-\lambda$. Therefore, the estimated confidence regions are computed as:

$$\left(A^C_{j\alpha} - \hat{A}^C_{j\alpha}\right)^2 + \left(A^W_{j\alpha} - \hat{A}^W_{j\alpha}\right)^2 < t^2_{j\alpha} S^2_{\hat{A}_{j\alpha}},\ j = 0, 1 \qquad (3)$$

where, $A^c_{j\alpha}$ and $A^w_{j\alpha}$ are the midpoint and width of $A_{j\alpha}$, j = 0, 1 respectively.

RESULTS AND DISCUSION

Experimentation of the model on a clinical data set: In this study, we use a numerical example to illustrate the fuzzy logistic model that discussed in previous sections. The data for this application was taken from (Namdari *et al*., 2015), Systematic Lupus Erythematosus (SLE) is a chronic autoimmune disease which attacks multiple systems in the body including the skin, blood, lungs, heart, brain and nervous system. There is no single diagnostic test for SLE. Physicians have to use a list of 11 criteria to help them in the diagnosis of SLE. Generally, a person needs to satisfy at least 4 out of the 11 criteria before a diagnosis can be made. The question which arises here is about a person with 3 symptoms. Is he/she considered as a healthy case without any therapy? Or is the severity of disease the same in patients with 4, 5, ..., 11 criteria? So, the distinction borderline between patients and healthy people cannot be considered crisp in SLE disease. Therefore, an expert was asked to assign a possibility of disease to each case using linguistic terms; very low, low, medium, high and very high (Namdari *et al*., 2015). Our sample contains 15 females aged 18-40 who are suspected to have SLE. Anti Nuclear Antibody (ANA) test that is a diagnostic blood test is used as the significant risk factor (X) (Table 1).

Table 1: The sample data

Possibility of SLE	X
High	112
Medium	80
High	115
High	105
Medium	89
Very high	160
Medium	100
High	100
Low	48
Very low	15
Low	50
Medium	59
Low	83
Low	15
Medium	85

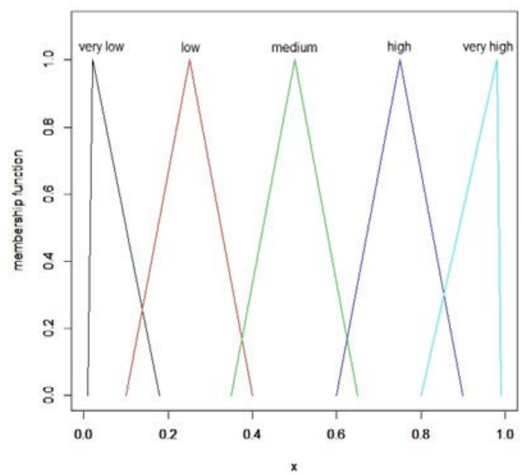

Fig. 1: Possibilities of SLE

For this situation, Namdari *et al*. (2015) assume the observations as triangular fuzzy numbers that cover the range of (0, 1) (Fig. 1). To define the relationship between possibility odds of SLE disease and the mentioned risk factor we use the following model:

$$\hat{y}_i = \log\frac{\tilde{p}_i}{1-\tilde{p}_i} = \tilde{A}_0 + \tilde{A}_1 x_i + \tilde{\varepsilon}_i, j = 1, 2, ..., 15$$

The lower and upper bounds of the α-cuts of $\hat{A}_0$ and $\hat{A}_1$ for α = 0, 0.1, 0, ..., 0.9, 1 are provided in Table 2. α = 1.0 shows the regression coefficient that is most likely and the α = 0.0 shows the range in which the regression coefficient could appear. In this example, although these two regression coefficients $\hat{A}_0$ and $\hat{A}_1$ are fuzzy, their most likely values are -3.181 and 0.0392, respectively and it is impossible for their values to take outside the ranges of [-3.732, -2.032] and [0.0349, 0.0360], respectively. For case 1, as an example with α = 0 the α-cut of estimated output based on our proposed model is:

Table 2: Estimated α-cuts of $\tilde{A}_0$ and $\tilde{A}_1$ 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	1.0
A_0^L	-3.732	-3.599	-3.485	-3.385	-3.300	-3.228	-3.169	-3.126	-3.102	-3.109	-3.181
A_0^U	-2.032	-2.089	-2.155	-2.228	-2.310	-2.401	-2.503	-2.621	-2.760	-2.933	-3.181
A_1^L	0.0349	0.0345	0.0342	0.0340	0.0340	0.0341	0.0345	0.0350	0.0357	0.0370	0.0392
A_1^L	0.0360	0.0356	0.0353	0.0351	0.0351	0.0352	0.0354	0.0358	0.0365	0.0375	0.0392

Table 3: 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	0.614	0.592	0.566	0.535	0.506	0.473	0.439	0.374	0.283	0.133
Pourahmad *et al*. (2011a, b) model	0.526	0.498	0.470	0.437	0.340	0.356	0.303	0.244	0.198	0.097

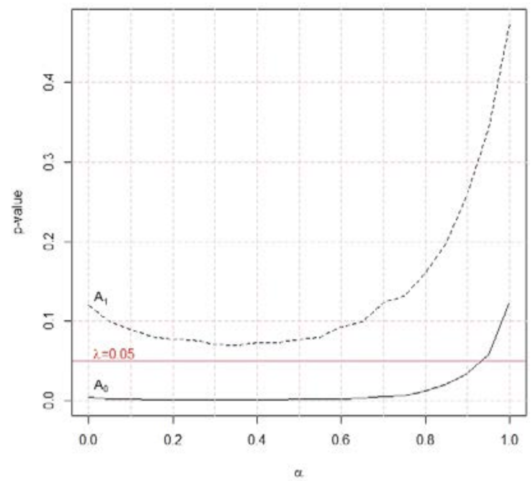

Fig. 2: The Behavior of p-value

$$\hat{y}_{1\alpha} = [-3.732, -2.032] + [0.0349, 0.0360] \times 112$$

So that:

$$\hat{y}_{1\alpha} = \left(\log \frac{\tilde{p}_1}{1-\tilde{p}_1}\right)_\alpha = [0.177, 2.000]$$

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

$$\hat{p}_{1\alpha} = \left[\frac{\exp(0.177)}{1+\exp(0.177)}, \frac{\exp(2.000)}{1+\exp(2.000)}\right] = [0.544, 0.881],\ \alpha = 0$$

The observed α-cut of possibility of SLE for this case for α = 0 is [0.6,0.9]. Now consider a new case with the ANA test equals 110. The estimated α-cut (for α = 0) of output for this new case using our proposed model is:

$$\hat{y}_{0\alpha} = [-3.732, -2.032] + [0.0349, 0.0360] \times 110 = [0.107, 1.928]$$

So, the estimated α-cut of possibility of disease for this case is: $p_{0\alpha}$ = [0.527, 0.873], α = 0 To evaluate the model based on the proposed index (Definition 4) we calculate MCIα for various α∊[0,1] and compare our model with the model proposed by Pourahmad *et al*. (2011a, b). As in Table 3, the MCIα index of our model is greater than other's for all α's. Also, the mean of this index for a sequence of 1000 's in our model is 0.493 and in other model is 0.384. To perform the test, 10000 replicate data sets were created by bootstrap method using the residuals. We consider two hypotheses as follows:

$$H_0: A_{j\alpha} = A^*_{j\alpha} \text{ vs } H_1: A_{j\alpha} \neq A_{j\alpha},\ j = 0, 1$$

where, $A_{0\alpha}$ and $A^*_{1\alpha}$ are the α-cuts of two symmetric triangular fuzzy numbers with centers -3.859 and 0.0431, respectively and spreads 1.162 and 0, respectively. These fuzzy numbers are estimated parameters of the model proposed by Pourahmad *et al*. (2011a). For each α∊{0.0, 0.1, ..., 0.9, 1.0}. Table 4 and 5 show test statistic and p-value of hypothesis test about $A_{0\alpha}$ and $A_{1\alpha}$, respectively.

If the p-value is smaller than significance level λ,then the null hypothesis is rejected. Figure 2 shows the behavior of the p-value for various α-cuts of each A_j, j = 0, 1. As shown in Fig. 2, for the intercept when α-cuts are >0.93 we accept the null hypothesis at the level of significance λ = 0.05 and when α-cuts are <0.93 we reject the null hypothesis. Also for the slope we accept the null hypothesis at the level of significance λ = 0.05 for all α. This example shows that the statistical significance of the coefficients changes depending on the vagueness of the data.

Also, the hypotheses H_0: $A_{j\alpha}$ = {0} vs H_1: $A_{j\alpha} \neq$ {0}, j = 0, 1 were tested by the mentioned bootstrap method and obtained p-values for various 's were equal to zero. Therefore both slope and intercept are strongly significant at λ level. The 95% confidence region defined in 4 for A_0, A_1 is provided in Fig. 3 where the x-axis denotes the center $A^c_{j\alpha}$ = $A^U_{j\alpha}$+$A^L_{j\alpha}$/2 of $A_{j\alpha}$ = [$A^l_{j\alpha}$, $A^u_{j\alpha}$] the y-axis represents the width $A_{j\alpha}$ = [$A^L_{j\alpha}$+$A^U_{j\alpha}$] of and z-axis is α.

Table 4: Test statistics and p-values for $A_{0\alpha}$

α	0.000	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900	1.000
$T_{0\alpha}$	5.3890	5.6610	5.8390	5.9310	5.9440	5.8770	5.7210	5.4590	5.0510	4.4150	3.3310
p-value	0.0037	0.0018	0.0011	0.0005	0.0012	0.0016	0.0020	0.0053	0.0120	0.0343	0.1262

Table 5: Test statistics and p-values for $A_{1\alpha}$

α	0.000	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900	1.000
$T_{1\alpha}$	1.5400	1.6680	1.7560	1.8080	1.8240	1.8040	1.7430	1.6340	1.4600	1.8730	0.7290
p-value	0.1197	0.0897	0.0769	0.0707	0.0726	0.0775	0.0934	0.1232	0.1611	0.2595	0.4730

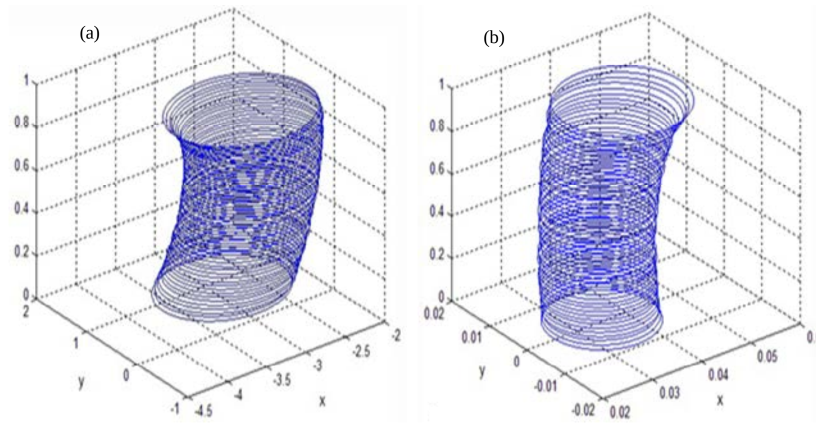

Fig. 3: The 95% confidence regions of: a) A_0 and b) A_1

CONCLUSION

In this study, researchers 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. Also, it sounds the introduced method is simpler and more accurate than that of Pourahmad *et al.* (2011a, b) method.

The proposed model is recommended for crisp input and fuzzy binary output observations. This adaptive model uses the logarithmic transformation of possibility of success (pi) for each case. We consider pi 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.

Researchers 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. At the end, a numerical example in the clinical field was used to detect the applied aspect of our model.

RECOMMENDATIONS

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.

REFERENCES

Akbari, M.G., R. Mohammadalizadeh and M. Rezaei, 2012. Bootstrap statistical inference about the regression coefficients based on fuzzy data. Int. J. Fuzzy Syst., 14: 549-556.

Celmins, A., 1987a. Least squares model fitting to fuzzy vector data. Fuzzy Sets Syst., 22: 245-269.

Celmins, A., 1987b. Multidimensional least-squares fitting of fuzzy models. Math. Modelling, 9: 669-690.

Chen, Y. and L. Wei, 2016. Parameter estimation of the fuzzy logistic regressive model with LR typed fuzzy coefficients. Oper. Res. Fuzziol., 6: 27-35.

Diamond, P., 1988. Fuzzy least squares. Inform. Sci., 46: 141-157.

Efron, B., 1979. Bootstrap methods: Another look at the Jackknife?. Annal. Statistics, 7: 1-26.

Hauser, D.R.A., D.T.F. Oliveira, D.A.M. Silveira, P.J.M. Brazao and R.L. Ziolli, 2012. Logistic regression and fuzzy logic as a classification method for feral fish sampling sites. Environ. Ecol. Statistics, 19: 473-483.

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

Nagar, P. and S. Srivastava, 2008. Adaptive fuzzy regression model for the prediction of dichotomous response variables using cancer data: A case study. J. Appl. Math. Stat. Inf., 4: 183-191.

Namdari, M., J.H. Yoon, A. Abadi, S.M. Taheri and S.H. Choi, 2015. Fuzzy logistic regression with least absolute deviations estimators. Soft Comput., 19: 909-917.

Phibanchon, S., S.A. Kareem, R. Zain and B. Abidin, 2007. An adaptive fuzzy regression model for the prediction of dichotomous response variables. Proceedings of the International Conference on Computational Science and its Applications (ICCSA07), August 26-29, 2007, IEEE, New York, USA., ISBN:0-7695-2945-3, pp: 14-19.

Pourahmad, S., S.M.T. Ayatollahi and S.M. Taheri, 2011a. Logistic regression, a new posssibilistic regression and its application in clinical vague status. Iran. J. Fuzzy Syst., 8: 1-17.

Pourahmad, S., S.M.T. Ayatollahi, S.M. Taheri and Z.H. Agahi, 2011b. Fuzzy logistic regression based on the least squares approach with application in clinical studies. Comput. Math. Appl., 62: 3353-3365.

Wu, C.F.J., 1986. Jackknife, bootstrap and other resampling methods in regression analysis. Ann. Stat., 14: 1261-1295.

Zadeh L.A., 1975a. The concept of linguistic variable and its application to approximate reasoning. Inf. Sci., 8: 199-249.

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