A NEW APPROACH FOR PARAMETER ESTIMATION IN FUZZY LOGISTIC REGRESSION

G. ATALIK AND S. SENTURK

Abstract. Logistic regression analysis is used to model categorical dependent variable. It is usually used in social sciences and clinical research. Human thoughts and disease diagnosis in clinical research contain vagueness. This situation leads researchers to combine fuzzy set and statistical theories. Fuzzy logistic regression analysis is one of the outcomes of this combination and it is used in situations where the classical logistic regression assumptions' are not satisfied. Also it can be used if the observations or their relations are vague. In this study, a model called Fuzzy Logistic Regression Based on Revised Tanaka's Fuzzy Linear Regression Model is proposed. In this regard, the methodology and formulation of the proposed model is explained in detail and the revised Tanaka's regression model is used to estimate the parameters. The Revised Tanaka's Regression model is an extension of Tanaka's Regression Model in which the objection function is developed. An application is performed on birth weight data set. Also, an application of diabetes data set used in Pourahmad et al.'s study was conducted via our proposed data set. The validity of the model is shown by the help of goodness of fit criteria called Mean Degree Memberships (MDM).

1. Introduction

The logistic regression model is the most frequently used regression model in situations where the dependent variable is categorical. Logistic regression is preferred, when the concerned dependent variable is composed of categories such as "patient - not patient", "positive negative"[6].

Classic (two valued) logic that first started with Aristotle turned into three valued logic, then multi-valued logic and eventually fuzzy logic with the progress of science and logic systems for the centuries. Fuzzy logic is a kind of logic including all logic systems [1].

It is difficult to model the systems in which, human thought and experience is effective. Because, those systems have ambiguous structure (pattern). Several theories are used to understand and determine the uncertainty. Probability theory and statistical methods are used commonly to model uncertainty. However, all of the vagueness confronted in daily life cannot be explained by randomness. Using statistics and probability theories that consider numerical vagueness may not be the correct way to identify non-random uncertainties. The fuzzy set theory proposed

Received: February 2016; Revised: February 2017; Accepted: May 2017

Key words and phrases: Fuzzy logistic regression, Revised Tanaka regression model, MDM criteria.

by Zadeh in 60's [16] is more successful at identifying these uncertainties. Fuzzy set theory allows us to evaluate qualitative variables as numerical [1].

Fuzzy logistic regression analysis is a regression method, based on fuzzy set theory, used in situations where the assumptions of classical logistic regression analysis cannot be fulfilled or are fuzzy due to the nature of data [13]. Moreover, it can be used when the observations or their relations are vague.

There are a few studies about fuzzy logistic regression in literature. The papers about this topic are shown below.

Nagar ve Srivastava's [10] paper presented an adaptive technique for the forecasting of binary dependent variable. This technique was a combination of fuzzy logic and statistical logistic regression. Their adaptive fuzzy logistic regression model was based on Tanaka's possibilistic regression analysis. In this model, data needed to be transformed. They tested the model on a cancer data set.

Dom et. al. [3] combined the classical regression method with the fuzzy concept. This model was also based on Tanaka's possibilistic regression analysis.They suggested a new learning algorithm for prediction of binary response variable. Linear transformation and fuzzy regression were included in the proposed algorithm.

Pourahmad et al. [13] introduced a new term called "possibilistic odds" due to the incompatibility of using logistic regression when used in clinical research for some disease diagnosis. They also introduced "fuzzy logistic regression" based on possibilistic odds. They mentioned coefficient estimation and goodness of fit tests for fuzzy logistic regression. The model was tested on a diabetes data set.

Pourahmad et al. [14] proposed an approach based on the least squares method for prediction of fuzzy logistic regression's coefficients. The model was performed with Lupus Erythematosus data. A new goodness of fit criteria called "Capability index" was also introduced.

Namdari et al. [11] used fuzzy logistic and ordinal logistic regressions to examine the effect of folic acid on appetites in children. They compared the fuzzy logistic regression model with a statistical ordinal logistic regression model and interpreted the results.

Namdari et. al. [12] proposed a new estimator for fuzzy logistic regression. This estimator was called "Least Absolute Deviations". They gave two numerical example. In their paper, two new goodness - of fit indices called measure of performance based on fuzzy distance and index of sensitivity were introduced and the results were compared with the least squares estimation method.

The aim of this paper is to examine fuzzy logistic regression analysis in detail and to propose a new approach to estimate model parameters. The estimation of model parameters were obtained by Tanaka's fuzzy linear regression model (FLR) in the fuzzy logistic regression method proposed by Pourahmad et al. [13]. He et al.[4] revised Tanaka's FLR method in their studies and proposed a more effective method than Tanaka's FLR method. In this study, the estimation of fuzzy logistic regression model's coefficient is conducted by Revised Tanaka's FLR as proposed by He et al.[4]. The theoretical structure of "Fuzzy Logistic Regression Based on Revised Tanaka FLR" and its applications are discussed for the first time in this study.

2. **Fuzzy Logistic Regression Analysis**

Regression analysis is a statistical technique that investigate the relationship between dependent variable and independent variables and models this relationship mathematically. The variables we are establishing a relationship between must be numeric in this analysis. Regression analysis can be used, when the some assumptions hold true. These assumptions are; errors are distributed normal with zero mean and constant variance and don't have relation (autocorrelation) [9].

The most important feature that distinguishes logistic regression from classic regression is that, dependent variable in logistic regression is categorical. The difference between the analyses reflects on both the selection of parametric models and assumptions. Logistic regression tries to make an estimation based on some variable values as in linear regression analysis. There are differences between the logistic regression and linear regression. The dependent variable of logistic regression is categorical. The estimation of the values of dependent variable is calculated by possible occurrence in logistic regression. Also, there is any assumptions on dependent variable's distributions [7].

Surveys on social sciences and marketing which use the field of logistic regression contain human thought and emotions. Human thought and emotions have a vague structure. Similarly, linguistic terms like "low, medium, high" are used in clinical research to measure the asperity of disease in patients. These linguistic terms can be seen as fuzzy representations of a person's opinion. Also, despite measuring with a numerical scale the boundary between sets is not crisp (1: low, 2: medium etc.) For example, an oral glucose tolerance test is used to determine diabetic patients. The cut-off point of test is 140 *[milligram (mg)]/[decilitre (dl)]* for two-hour plasma glucose. This value is not a precise borderline [13].

Cases in the neighborhood of borderline may be fuzzy depending on illness. Statistical methods based on crisp observation and certain assumptions are insufficient to model the relationship between such observations. In practice, there are many situations in which the logistic regression method, which models the relationship between categorical dependent variable and independent variables, cannot be used. For instance, the absence of applicable instruments or distinct and wholly accepted criteria for some illness in clinical research causes researchers to encounter vague observations [13].

The observed random variable contains variability or fuzziness from the source of uncertainty. Given that the random variable includes uncertainty, it can be determined by fuzzy models, while the determination can be made by stochastic models in case of a variability occurrence. However, it contains both, in that case it can be identified with fuzzy statistics models.

Fuzzy logic theory and statistics theory are complementary theories [17]. Techniques composed of these theories succeed in identifying and modelling data in case where the assumptions of statistical techniques do not hold true or statistical methods are not proper. As a result of these reasons, "Fuzzy Logistic Regression Analysis" has emerged as an integration of fuzzy regression and logistic regression.

Fuzzy logistic regression analysis is a fuzzy approach to model the dependent variable and it is used in cases not holding the assumptions of logistic regression and arising imprecise data. (In cases where the assumption of logistic regression are do not hold true and there is imprecise data.)

Logistic regression analysis is based on certain assumptions like other statistical methods and these assumptions sometimes lead to the emergence of some problems in practice. As an illustration, diagnosis are doubted by doctors due to the absence of sufficient equipment or well-defined criteria in clinical research based on a set of independent variables. Thus, doctors cannot classify an individual in one of two categories. In these circumstances, binary dependent observations are vague and the relationship between varibles is not definite to implement logistic regression. The probability of one's belonging to category 1 ($p = P(Y = 1)$) and the odds ratio ($p/(1-p)$) are not calculated by virtue of the ambiguity of the dependent variable. Usually, such observations are not included in the model in logistic regression analysis. Supposing all observations are vague, the relationship between the probability of success and independent variables is not modelled by classical statistical approaches. This kind of uncertainty is not arisen by randomness and probability. It is necessary to consider the other kind of uncertainty such as the possibility to model such observations. Each observation is compared to formerly confirmed (accepted) criteria of class 1 members and the possibility of belonging to class 1 is noted by conferring with a professional. So,"possibilistic odds" is specified and modeled.

Supposing that the possibility of belonging to a class with known characteristic for each vague observation is represented by μ_i and its complement is shown by $1-\mu_i$ the ratio $\mu_i/(1-\mu_i)$ which demonstrates the possibility of having a considered feature for the i-th case to not is called the possibilistic odds [13].

Now that possibilistic odds which is needed to carry out fuzzy logistic regression is given, finally fuzzy logistic regression model can be obtained and it is shown in the coming section.

2.1. **The Theory of Proposed Model.** Suppose the data set $(x_{1i}, x_{2i}, \ldots, x_{ni})$ $i = 1, 2, \ldots, n$ where X_i is the vector of a crisp observation on the independent variable. Here, X_i could be both categorical and numerical variables. μ_i demonstrates the possibility of i-th case having the relating property for dependent variable. In other words, $\mu_i = Poss(Y_i = 1)$. Thus, fuzzy logistic regression model is described as follows:

$$\tilde{W}_i = \tilde{b_0} + \tilde{b_1} x_{1i} + \ldots + \tilde{b_n} x_{ni} \quad i = 1, 2, \ldots, n \tag{1}$$

$\tilde{b_0}, \tilde{b_1}, \ldots, \tilde{b_n}$ are triangular fuzzy numbers and model parameters. Also, $\tilde{W}_i = \ln(\mu_i/(1-\mu_i))$ is the estimator of the logarithmic transformation of possibilistic odds. $\tilde{b_j} = \left(b_j^c, s_j^L, s_j^R\right)_T, \quad j = 1, 2, \ldots, m$ is treated as triangular fuzzy number to simplify the calculation. Because of being model parameters triangular fuzzy numbers, fuzzy output $\tilde{W}_i$ is also triangular fuzzy number and it is indicated as: $\tilde{W}_i = \left(f_i^c(x), f_{is}^L(x), f_{is}^R\right)$. Where:

$$f_i^c(x) = a_0^c + a_1^c x_{1i} + \ldots + a_n^c x_{ni}$$

$$f_{is}^{L}(x) = s_0^L + s_1^L x_{1i} + \ldots + s_n^L x_{ni}$$
$$f_{is}^{R}(x) = s_0^R + s_1^R x_{1i} + \ldots + s_n^R x_{ni} \quad (2)$$

In the case of $s_i^L = s_i^R = s_i$, the triangular fuzzy number is called the symmetric triangular fuzzy number. So, it is taken as $f_{is}^{L}(x) = f_{is}^{R}(x) = f_i^c(x)$ for $\tilde{W}_i$. The brief definition of fuzzy logistic regression is given in this section so far. For more detailed theoretical information, see Pourahmad et al. [13].

The deviations between the observed and estimated values are supposed to arise from observation errors in the classical regression models. Tanaka et.al [15] assumed that these deviations depend on indefiniteness of the system structure. Because of that, they introduced FLR model. He et al.[4] clarified that the $h_i \quad (0 \leq h \leq 1)$ value, which is referred to as the degree of fit of the estimated fuzzy linear model to the given data, in Tanaka's FLR model depends not only on the estimated $\tilde{y}_i$'s spread, but also on the distance between $\tilde{y}_i$'s center and observed y_i. Therefore, they proposed a new model by developing the objection function in Tanaka's FLR. It is seen in the proposed model that, the system of fuzziness decreases and the average of estimated h_i values i.e $\bar{h}$ increases compared with Tanaka's model.

The novelty of this paper is to estimate the parameter of fuzzy logistic regression by the help of He et. al.'s [4] FLR which is mentioned in this paper Revised Tanaka FLR. In Pourahmad et. al.[13]'s paper, the parameter estimation of fuzzy logistic regression is based on Tanaka FLR's model. He et. al. [4] denoted that their method gives better estimation than Tanaka FLR's. So, He et. al's FLR is used in the parameter estimation of the fuzzy logistic regresion for the first time in this study. We give name to our method " Fuzzy Logistic Regression Based on Revised Tanaka FLR".

The parameters of fuzzy logistic regression can be estimated by using the objective function and constraints shown below.

$$\text{Minimizing:} \quad \text{SF} + \sum_{i=1}^{n} d_i = \sum_{i=1}^{n}(s_0 + \sum_{j=1}^{m} s_j|x_{ji}|) + \sum_{i=1}^{n} d_i \quad (3)$$

$$0 \leq h_i = 1 - d_i/(s_0 + \sum_{j=1}^{m} s_j|x_{ji}|)$$

$$d_i = |y_i - (c_0 + \sum_{j=1}^{m}(c_j x_{ji})|$$

$$i = 1, 2, \ldots, n \quad s_j \geq 0, \quad j = 0, 1, 2, \ldots, m \quad (4)$$

where SF is the system fuzziness, n is the number of observations y_i and x_{ji} ith observed values for y and x_j and h_i represents the membership for observed y_i belonging to estimated $\hat{\tilde{y}}$.

2.2. **Goodness - of - Fit Criteria for Fuzzy Logistic Regression Model.** Models based on fuzzy rules should be assessed by the methods as other statistical models. In other words, it should be checked whether the model fits the data or not. There are lots of goodness of fit criteria, but in this study it is mentioned about "Mean Degree of Memberships" for the evaluation of fuzzy logistic regression model.

w_i indicates the observed value of dependent variable for i-th case and $\tilde{W}_i$ also indicates estimated value of response variable for i-th case. The MDM is calculated as follows:

$$\text{MDM} = \frac{1}{n}\sum_{i=1}^{n}\tilde{W}_i(w_i) = \frac{1}{n}\exp((\tilde{W}_i)\frac{\mu_i}{1-\mu_i}) \tag{5}$$

MDM takes value between 0 and 1. Values near 1 indicates that model fits data set well.[13]

3. Application of Fuzzy Logistic Regression Model Based on Revised Tanaka Regression Model

In this section of the study, we have used an application of fuzzy logistic regression with a birth weight data set. Firstly, fuzzy logistic regression model coefficients for the relevant data are estimated by Pourahmad et. al.'s method. Then, the estimation of unknown parameters is conducted by Fuzzy Logistic Regression Based on Revised Tanaka Model which is given for the first time in this study. Goodness of fit criteria are calculated for both models and results are interpreted. Finally, the diabetes mellitus data set that was used in Pourahmad et al.'s study is modelled by Fuzzy Logistic Regression Based on Revised Tanaka's FLR to show the validity of the proposed method.

There are plenty factors that affect the birth weight of babies in data set used in this study. Mothers smoking, mother's socioeconomic status, maternal age, bleeding during pregnancy, mother's height, week of birth and mother's diet have an effect on birth weight according to Krimi and Pence [8], Hirve and Ganatra [5] and Bircan [2].

Birth weight, mother's age, mother's diet and smoking habbit of mother are identified as the undertaken independent variables by the help of above information in this study. Dependent variable is the baby's birth weight is normal or not according to undertaken independent variables.

It could be thought that classical logistic regression analysis is suitable in determining whether a baby's birth weight is normal or not. However, there are plenty factors that effect birth weight, as indicated above. The story of pregnancy alters from mother to mother. So, this situation leads to uncertainty in determining baby's birth weight. Moreover, there is a vagueness due to the nature of data and the relation between variables. Because of these reasons, fuzzy logistic regression analysis should be used to model data set.

To apply fuzzy logistic regression, possibilistic odds as indicated in section 2 should be calculated. For this reason, obstetricians were consulted in order to assess the independent variables and their opinion was taken as experts to determine the possibility of the dependent variable.

25 samples used in the analysis were taken from a hospital database in Kayseri. The variables are;

Y : Baby's birth weight is normal or not.

X_1 :Age of mother

X_2 :The birth weight of baby

No:	X_1(year)	X_2(gr)	X_3	X_4	μ_i	$\frac{\mu_i}{1-\mu_i}$	$W_i = \ln(\frac{\mu_i}{1-\mu_i})$
1	21	2400	1	0	0.03	0.0309	-3.4761
2	22	3200	1	0	0.05	0.0526	-2.9444
3	23	3400	0	0	0.08	0.0870	-2.4424
4	27	3300	1	1	0.10	0.1111	-2.1972
5	27	3500	1	1	0.12	0.1364	-1.9924
6	22	3500	1	0	0.14	0.1628	-1.8153
7	28	3600	0	0	0.16	0.1905	-1.6582
8	28	3600	1	1	0.19	0.2346	-1.4500
9	32	3500	1	1	0.20	0.2500	-1.3863
10	35	3500	1	0	0.22	0.2821	-1.2657
11	25	3500	1	0	0.25	0.3333	-1.0986
12	27	3800	0	0	0.28	0.3889	-0.9445
13	18	3800	1	0	0.30	0.4286	-0.8473
14	22	3750	1	1	0.33	0.4925	-0.7082
15	31	4000	0	0	0.38	0.6129	-0.4896
16	27	3500	1	0	0.42	0.7241	-0.3228
17	38	4000	1	0	0.45	0.8182	-0.2007
18	21	4000	0	1	0.48	0.9231	-0.0804
19	28	4000	1	0	0.52	1.0833	0.0800
20	29	4250	1	1	0.59	1.4390	0.3640
21	20	4400	1	0	0.75	3.0000	1.0986
22	28	4500	1	0	0.79	3.7619	1.3249
23	27	4500	0	0	0.81	4.2632	1.4500
24	30	5000	1	0	0.91	10.1111	2.3136
25	25	5500	1	0	0.95	19.0000	2.9444

TABLE 1. The Data Sets About Babies Birth Weight

X_3 :The smoking habit of mother. The mothers who don't smoke are coded as 0 and mothers who smoke are coded as 1.

X_4 :Mother's diet. The mothers who pay attention to their nutrition are coded as 1 and mothers who dont are coded as 0.

Concerned variables and their possibilistic odds are shown in Table 1.

The fuzzy logistic regression model that was used in the application is a model with precise input and fuzzy output and its theoretical structure is shown below.

$$\tilde{W}_i = \tilde{b}_0 + \tilde{b}_1 x_{1i} + \tilde{b}_2 x_{2i} + \tilde{b}_3 x_{3i} + \tilde{b}_4 x_{4i} \qquad (6)$$
$$\tilde{b}_0 = (b_0^c, s_0), \tilde{b}_1 = (b_1^c, s_1), \ldots, \tilde{b}_4 = (b_4^c, s_4)$$

The objective function in equation (7) must be solved according to constraints in equation (8) to estimate the parameters of fuzzy regression model shown in equation (6) based on the approach proposed by Pourahmad et al. [13]

$$Z = n(s_0^L + s_0^R) + \sum_{j=1}^{m}[(s_j^L + s_j^R)\sum_{i=1}^{n} x_{ji}] \qquad (7)$$
$$1 - \frac{f_i^c(x) - w_i}{f_{is}^L(x)} \geq h \Rightarrow (1-h)s_0^L + (1-h)\sum_{j=1}^{m} s_j^L x_{ij} - a_0^c - \sum_{j=1}^{n} a_j^c x_{ij} \geq -w_i$$

$$1-\frac{w_i-f_i^c(x)}{f_{is}^L(x)} \geq h \Rightarrow (1-h)s_0^R + (1-h)\sum_{j=1}^{m} s_j^R x_{ij} + a_0^c + \sum_{j=1}^{m} a_j^c x_{ij} \geq w_i \tag{8}$$

The objective function in equation (7) can be calculated as the following for the data sets shown in Table 1.

$$Z = 2(25s_0 + s_1\sum_{i=1}^{25} X_{1i} + s_2\sum_{i=1}^{25} X_{2i} + s_3\sum_{i=1}^{25} X_{1i} + s_4\sum_{i=1}^{25} X_{4i})$$

$$= 2(25s_0 + 661s_1 + 96000s_2 + 19s_3 + 7s_4) \tag{9}$$

The objective function shown above must be solved under 50 constraints. The "h" term draws the attention in equation (8). The "h" term is decided by decision - makers before analysis. The two constraints for $h = 0.5$ for the first and the last data in Table 1 are as follows: For the first data;

$$c_0+21\times c_1+2400\times c_2+1\times c_3+0\times c_4-0.5\times s_0-0.5\times 21\times s_1-0.5\times 2400\times s_2-0.5\times 1\times s_3$$
$$-0.5\times 0\times s_4 \leq -3.4761$$
$$c_0+21\times c_1+2400\times c_2+1\times c_3+0\times c_4+0.5\times s_0+0.5\times 21\times s_1+0.5\times 2400\times s_2+0.5\times 1\times s_3$$
$$+0.5\times 0\times s_4 \geq -3.4761 \tag{10}$$

For the last data;

$$c_0+25\times c_1+5500\times c_2+1\times c_3+0\times c_4-0.5\times s_0-0.5\times 25\times s_1-0.5\times 5500\times s_2-0.5\times 1\times s_3$$
$$-0.5\times 0\times s_4 \leq 2.9444$$
$$c_0+25\times c_1+5500\times c_2+1\times c_3+0\times c_4+0.5\times s_0+0.5\times 25\times s_1+0.5\times 5500\times s_2+0.5\times 1\times s_3$$
$$+0.5\times 0\times s_4 \geq 2.9444 \tag{11}$$

When h is fixed as 0.5, the objective function is $Z = 213.285$ and the model is shown below.

$$\tilde{W}_i = (0.0000, 2.1674) + (-0.1652, 0.0556)X_1 + 0.0009X_2$$
$$+(-0.1000, 0.8275)X_3 + 0.3395X_4 \tag{12}$$

We can calculate the possibilistic odd of 8. observation in Table 1 via equation (12).

$$\tilde{W}_i = (0.0000, 2.1674) + (-0.1652, 0.0556)\times 28 + 0.0009\times 3600$$
$$+(-0.1000, 0.8275)\times 1 + 0.3395\times 1 = (0.0000, 2.1674) + (-4.6256, 1.5568)$$
$$(3.2400, 0.0000) + (-0.1000, 0.8275) + (0.3395, 0.0000) = (-1.1461, 4.5517) \tag{13}$$

After obtaining the results of the above equation, now we can calculate the estimated possibilistic odds of 8. observation using extension principle.

$$exp(\tilde{W}_8(x)) = \begin{cases} 1-\frac{-1.15-\ln(x)}{4.55} & -5.7 \leq \ln(x) \leq -1.15 (0.0033 \leq x \leq 0.32) \\ 1-\frac{\ln(x)+1.15}{4.55} & -1.15 \leq \ln(x) \leq 3.41 (0.32 \leq x \leq 30.13) \end{cases} \tag{14}$$

The possibilistic odds of 8. observation is founded as 0.32. This value shows, the ratio of the possibility that the baby's birth weight is normal against the possibility that the baby's birth weight is not normal. Related reverse transformation is done, we can say the possibility that this baby has normal birth weight is 0.24.

We can calculate a new person's possibilistic odds. Suppose that, there is a new person with $(x_1 = 24, x_2 = 4500, x_3 = 1, x_4 = 0)$. According to model shown in equation (13), the possibilistic odds for this new person is calculated as follows:

$$\tilde{W}_{new} = (0.0000, 2.1674) + (-0.1652, 0.0556) \times 24 + 0.0009 \times 4500$$
$$+(-0.1000, 0.8275) \times 1 + 0.3395 \times 0 = (0.0000, 2.1674) + (-3.9648, 1.3344)$$
$$(4.0500, 0.0000) + (-0.1000, 0.8275) + (0.0000, 0.0000) = (-0.0148, 4.3293) \quad (15)$$

To make calculations easier, the $\tilde{W}_{new}$ value is taken $\tilde{W}_{new} = (-0.015, 4.33)$. The possibilistic odds for new person can be calculated as follows:

$$exp(\tilde{W}_{new}(x)) = \begin{cases} 1 - \frac{-0.015 - \ln(x)}{4.33} & -4.345 \leq \ln(x) \leq -0.015 (0.013 \leq x \leq 0.98) \\ 1 - \frac{\ln(x) + 0.015}{4.33} & -0.015 \leq \ln(x) \leq 4.315 (0.98 \leq x \leq 74.81) \end{cases} \quad (16)$$

So, we can say that, the possibilistic odds of the newborn is about 0.98. After doing relevant reverse transformations, we can say that, the possibility of new person's birth weight is normal is 0.50.

Mean Degree of Memberships described in the previous section is a criterion similar to coefficient of determination in classical regression and indicates how well the constructed model fits to the data set. The mean degree of memberships is calculated for the model above as follows:

$$MDM = \frac{1}{n} \sum_{i=1}^{n} \tilde{W}_i(w_i) = \frac{1}{25} \times 17.96 \Rightarrow MDM = 0.7210 \quad (17)$$

MDM value is calculated as 0.7210. This value can be interpreted as the fuzzy logistic regression model indicates a good fit to the data set.

The prediction of fuzzy logistic regression model's coefficients will be performed by Fuzzy Logistic Regression based Revised Tanaka regression model proposed in this study with the same data set.

The objective function in equation (19) must be solved according to constraints in equation (20) to estimate the parameters of fuzzy logistic regression model shown in equation (18)

$$\tilde{W}_i = \tilde{b}_0 + \tilde{b}_1 x_{1i} + \tilde{b}_2 x_{2i} + \tilde{b}_3 x_{3i} + \tilde{b}_4 x_{4i} \quad i = 1, 2 \ldots, 25 \quad (18)$$

$$Z = \text{SF} + d_i = \sum_{i=1}^{n} (s_0 + \sum_{j=1}^{m} s_j |x_{ji}|) + \sum_{i=1}^{n} d_i \quad (19)$$
$$0 \leq h_i = 1 - d_i \Big/ (s_0 + \sum_{j=1}^{m} s_j |x_{ji}|)$$

$$d_i = |y_i - (c_0 + \sum_{j=1}^{m} c_j x_{ji})| \quad i = 1, 2, \ldots, n \quad (20)$$

	(c_0, s_0)	(c_1, s_1)	(c_2, s_2)	(c_3, s_3)	(c_4, s_4)	**Z**
Model I	(0.0 , 2.167)	(-0.165 , 0.05)	(-0.0009 , 0.0)	(-0.10 , 0.827)	(0. 339 , 0.0)	213.285
Model II	(0.0 , 0.468)	(-0.078 , 0.0)	(0.00039 , 0.001)	(-0.823 , 0.972)	(0.1440 , 0.0)	288.053

TABLE 2. Model Parameters and Objective Functions for $h = 0.5$

The objective function in equation (19) is calculated as follows:

$$Z = 2(25s_0 + s_1 \sum_{i=1}^{25} X_{1i} + s_2 \sum_{i=1}^{25} X_{2i} + s_3 \sum_{i=1}^{25} X_{3i} + s_4 \sum_{i=1}^{25} X_{4i}) + \sum_{i=1}^{25} d_i$$
$$= 2(25s_0 + 661s_1 + 96000s_2 + 19s_3 + 7s_4) + d_1 + d_2 + \ldots + d_{25} \quad (21)$$

Above objective functions' constraints are shown in equation (22) and equation (23) for the first and last data. For the first data;

$$d_1/(s_0 + 21s_1 + 2400s_2 + s_3) \leq 0.5$$
$$d_1 = |-3.4761 - c_0 - 21c_1 - 2400c_2 - c_3| \quad (22)$$

For the last data;

$$d_{25}/(s_0 + 25s_1 + 5500s_2 + s_3) \leq 0.5$$
$$d_{25} = |2.9444 - c_0 - 25c_1 - 5500c_2 - c_3| \quad (23)$$

The coefficients obtained from Lingo 14. software are as follows: $c_0 = 0.0000$, $s_0 = 0.4684$, $c_1 = -0.0783$, $s_1 = 0.0000$, $c_2 = 0.00039$, $s_2 = 0.00103$, $c_3 = -0.8238$, $s_3 = 0.9721$, $c_4 = 0.1440$, $s_4 = 0.0000$. The objective function is calculated as $Z = 288.053$ for the fuzzy logistic regression model shown in equation (24)

$$\tilde{W}_i = (0.0000, 0.4684) + -0.0783X_1 + (0.00039, 0.00103)X_2$$
$$+(-0.8238, 0.9721)X_3 + 0.1140X_4 \quad (24)$$

After the fitting model, we can calculate the goodness of fit criteria. The mean degree of memberships is calculated for Fuzzy logistic regression based on Revised Tanaka Model as shown below.

$$MDM = \frac{1}{n} \sum_{i=1}^{n} \tilde{W}_i(w_i) = \frac{1}{25} \times 19.47 \Rightarrow MDM = 0.7813 \quad (25)$$

The MDM criteria is calculated as 0.7813 for the proposed model. This value indicates a good fit and it can be interpreted that the fuzzy logistic regression model is good at modelling the birth weight data set.

The parameters of fuzzy logistic regression are obtained by Pourahmad et al's approach and the method proposed in this study for different h values. Model I shows the model proposed by Pourahmad et al. and Model II shows the model proposed in this study. The coefficients, objective functions for these models are shown for $h = 0.5$ and $h = 0.7$ in Table 2, Table 3. Also MDM criteria for all h values between 0.1 and 0.9 is shown in Table 4.

As it can be seen in Table 4, the MDM criteria for the model proposed in this study is higher at all h levels. According to MDM criteria, The Fuzzy Logistic Regression Model Based on Revised Tanaka Model gives better results.

	(c_0, s_0)	(c_1, s_1)	(c_2, s_2)	(c_3, s_3)	(c_4, s_4)	**Z**
Model I	(0.0 , 1.548)	(-0.165 , 0.039)	(0.0009 , 0.0)	(-0.10 , 0.591)	(0. 339 ,0.0)	152.346
Model II	(0.0 , 1.380)	(-0.0733 , 0.0)	(0.0003 , 0.0004)	(-0.633 , 0.692)	(0.056 , 0.0)	214.665

TABLE 3. Model Parameters and Objective Functions for $h = 0.7$

$h =$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Model I	0.9014	0.8884	0.8326	0.7731	0.7210	0.6651	0.6094	0.5534	0.4979
Model II	0.9558	0.9098	0.8671	0.9885	0.7813	0.9268	0.7086	0.9692	0.6276

TABLE 4. The MDM Criteria for All h Values

4. **Conclusion**

Logistic regression analysis depends on some assumptions. Sometimes these assumptions cannot be fulfilled due to vagueness in the data set. If the undertaken data set involves vagueness, classical statistical methods cannot be applied. Fuzzy statistic theory which is a combination of fuzzy set and classical statistical theories can be used in this situation. Fuzzy logistic regression analysis is a method to model the dependent variable in cases where the data set contains vagueness or the model assumptions of logistic regression are not satisfied.

In this paper, Fuzzy Logistic Regression Based on Revised Tanaka's FLR is introduced for the first time. We used the Revised Tanaka's FLR to estimate the model parameter. A practical application of our model is demonstrated using a babies birth weight data set. Mean Degree of Memberships is used to determine the success of the model. It is seen that via the MDM criteria, our proposed model gives better result than Pourahmad et al's method [13]. The possible reason for that is we used revised Tanaka FLR model to estimate the model parameter which gives better estimation than Tanaka's FLR.

The proposed model can be used in case of vagueness in the data set. Fuzzy Logistic Regression Based on Revised Tanaka's FLR model can be seen an alternative to Pourahmad et. als model. For furthere research, both models can be used and compared in other similar research areas and the different types of fuzzy numbers can be used in the proposed model.

Acknowledgements. This work was supported by Anadolu University Scientific Research Projects (Project Number:1307F285)

REFERENCES

[1] G. Atalik, *A New Approach for Parameter Estimation in Fuzzy Logistic Regression and an Application*, Master of Science Thesis, Anadolu University, Graduate School of Sciences, Eskisehir (2014).

[2] H. Bircan, *Lojistik Regresyon Analizi: Tp Verileri zerine Bir Uygulama*, Kocaeli niversitesi Sosyal Bilimler Enstits Dergisi, **8(1)** (2004), 185-208.

[3] R. M. Dom, S. A. Kareem, A. Razak and B. Abidin, *A learning system prediction method using fuzzy regression*, In Proceedings of the International MultiConference of Engineers and Computer Scientists, Hong Kong, China, (2008), 19-21.

[4] Y. Q. He, L. K. Chan and M. L. Wu, *Balancing productivity and consumer satisfaction for profitability: statistical and fuzzy regression analysis*, European Journal of Operational Resarch, **176(1)** (2007), 252-263.
[5] S. S. Hirve and B. R. Ganatra, *Determinants of low birth weight: a commun,ty based prospective cohort study*, Indian Pediatrics, **31(10)** (1994), 1221-1225.
[6] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, John Wiley and Sons, New York, 2000.
[7] D. G. Kleinbaum and M. Klein, *Logistic Regression*, A Self-Learning Text (Second Edition ed.), Springer - Verlag , New York, 2002.
[8] E. Kirimi and S. Pence, *The affects of smoking during pregnancy to fetus and plasental development*, Van Medical Journal, **6(1)** (1999), 28-30.
[9] D. C. Montgomery, E. A. Peck and G. G. Vining, *Introduction to Linear Regression Analysis*, John Wiley and Sons, New York, 2001.
[10] P. Nagar and S. Srivastava, *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(2)** (2008), 183-191.
[11] M. Namdari, A. Abadi, S. M. Taheri, M. Rezaei, M. Kalantari and N. Omidvar, *Effect of folic acid on appetite in children: Ordinal logistic and fuzzy logistic regressions*, Nutrition, **30(3)** (2014), 274-278.
[12] M. Namdari, J. H. Yoon, A. Abadi, S. M. Taheri and S. H. Choi, *Fuzzy Logistic Regression with Least Absolute Deviations Estimators*, Soft Computing, **19(4)** (2015), 909-917.
[13] S. Pourahmad, S. M. T. Ayatollahi and S. M. Taheri, *Fuzzy logistic regression: a new possibilistic model and its application in clinical vague status*, Iranian Journal of Fuzzy Systems, **8(1)** (2011), 1-17.
[14] S. Pourahmad, S. M. Ayatollahi and S. M. Taheri, *Fuzzy logistic regression based on the least squares approach with application in clinical studies*, Computers and Mathematics with Applications, **62(9)** (2011), 3353-3365.
[15] H. Tanaka, S. Uejima and K. Asai, *Lineer regression analysis with fuzzy model*, IEEE Transactions On Systems, Man, and Cybernetics, **12(6)** (1982), 903-907.
[16] L. A. Zadeh, *Fuzzy sets*, Information and Control, **8** (1965), 338-353.
[17] L. A. Zadeh, *Discussion: probability theory and fuzzy logic are complementary rather than competitive*, Technometrics, **37(3)** (1995), 271-276.

Gultekin Atalik*, Department of Statistics, Anadolu University, Eskisehir, Turkey and Department of Statistics, Amasya University, Amasya,Turkey
E-mail address: gultekinatalik@anadolu.edu.tr, gultekin.atalik@amasya.edu.tr

Sevil Senturk, Department of Statistics, Anadolu University, Eskisehir, Turkey
E-mail address: sdeligoz@anadolu.edu.tr

*Corresponding author

A NEW APPROACH FOR PARAMETER ESTIMATION IN FUZZY LOGISTIC REGRESSION

G. ATALIK AND S. SENTURK

رویکرد جدید برای برآورد پارامتردر رگراسیون منطقی فازی

چکیده. آنالیز رگراسیون منطقی برای مدل سازی متغیر وابسته رسته ای و معمولاً در علوم اجتماعی و تحقیقات بالینی به کار برده می شود. تفکرات بشر و تشخیص بیماری در تحقیق بالینی شامل ابهامات است. این شرایط محققین را به سوی ترکیب مجموعه فازی و تئوری های آماری سوق می دهد. آنالیز رگرسیون منطقی یکی از پیامد های این ترکیب می باشد و در شرایطی به کار برده می شود که فرضیات رگرسیون منطقی کلاسیکی صادق نیستند. همچنین در صورتیکه مشاهدات یا روابط آنها مبهم باشند آنالیز رگرسیون منطقی می تواند به کار برده شود. در این تحقیق مدلی که رگرسیون منطقی فازی نامیده می شود و بر اساس مدل رگرسیون خطی فازی Tanaka باز بینی شده است ، پیشنهاد گردیده. در این خصوص متدولژی و فرمول بندی مدل پیشنهاد شده به طور کامل توضیح داده شده است و مدل رگرسیون باز بینی شده Tanaka جهت برآورد پارامترها به کار برده شده است . مدل رگرسیون بازبینی شده Tanaka توسیعی از مدل رگرسیون Tanaka است که درآن تابع هدف گسترش داده شده است. اجرا روی مجموعه وزن تولد کاربردی از آن است . همچنین به عنوان کاربرد دیگری از مجموعه داده های پیشنهادی ما اجرا روی مجموعه داده های دیابت است که توسط پور احمد و همکاران به کار برده شده است. درستی مدل توسط خوبی محک شایستگی که عضویت درجه میانگین نامیده می شود، نشان داده شده است.