International Journal of
Clinical Biostatistics and Biometrics

On the Comparison of Multiple Linear and Interval Regression Models of Apnea-Hypopnea Index in Patients with Obstructive Sleep Apnea Syndrome

Derviş Topuz[1]* iD *and Adnan Karaibrahimoğlu[2]* iD

[1]*Department of Health Services Science, Niğde Zübeyde Hanim Vocational School of Health Services, Niğde Ömer Halisdemir University, Turkey*

[2]*Department of Biostatistics and Medical Informatics, Medical School, Süleyman Demirel University, Turkey*

*Corresponding author: *Derviş Topuz, Department of Health Services Science, Niğde Zübeyde Hanım Vocational School of Health Services, Niğde Ömer Halisdemir University, 51200, Niğde, Turkey*

Abstract

Objective: In the fuzzy regression models that are fitted by using fuzzy logic, every regression coefficient should be estimated at a certain level of a fuzziness tolerance because of dividing the error value into the coefficient. This study aims to compute the coefficient and deviation values of classical least squares (OLS) and fuzzy interval regression models on a sample data set and interpret them comparatively.

Material and methods: There were totally 514 patients with OSAS (Obstructive Sleep Apnea Syndrome) who admitted to sleep outpatient clinics. The classical and interval regression models were established for 35 positional and non-positional patients who were selected randomly. It was suggested to use an h = 0.0 value called as "turbidity tolerance level", and the mean squared error (MSE) and the coefficient of determination (R^2) indexes were utilized as the goodness of fit test criteria showing the compatibility between the values calculated at the suggested h-level.

Results: According to the classical regression model with forward selection, R^2 values for dependent variable AHI (Apnea-Hypopnea Index) were calculated as $R^2_{non\text{-}positional}$ = 0.679 and $R^2_{non\text{-}positional}$ = 0.719, and the only independent variable ODI (oxygen desaturation index) was significant in both models. The turbidity (fuzziness) tolerance level of the model was calculated as h = 0.36 greater than h = 0.0. The fuzziness level of the interval regression models were calculated as 773.95 and 1129.54 respectively subject to 16 iterations and 70 constraints. R^2 values (0.98 and 0.91 respectively) were calculated as higher obtaining the predicted values, and MSE values were found as lower (11.08 and 5.94 respectively) than OLS models. For positional

OSAS patients ODI, REM and AH-duration had more significant effect on AHI whereas BMI, ODI and AROUSAL variables were strongly significant for non-positional OSAS patients.

Conclusions: It was shown that the independent variables Body Mass Index, Rapid Eye Movement, Oxygen Desaturation Index, Arousal (Sleep Stimulation) and Apnea-hypopnea duration had significant effect on AHI since their dispersions were very low in the interval regression models. It can be concluded that the results of the interval regression model were compatible with this dataset, and gave more accurate results.

Keywords

Fuzzy logic, Interval regression, Least square estimation, Apnea-Hypopnea index, Maximum dispersion

Introduction

Regression analysis is the statistical method that defines the causal relationship between a dependent variable and one or more independent variables, and are used to make relevant predictions [1]. The types and structures of dependent and independent data sets in the healthcare area (such as diagnosis, treatment, research, method development, etc.) often do not conform to classical models. Uncertainty situations in which the dependent variable cannot be expressed continuously or a set of independent variables could be affected by different environmental factors are encountered [2,3]. Moreover, there are many uncertain

situations such as the lack of suitable devices developed for measurement, the exact criteria for the classification procedures cannot be determined and the errors that may arise due to lack of sampling [4]. Shortly, assumptions for classical regression modeling are often difficult. In order to cope with such problems, the classical regression analysis method accepts errors between observed and predicted values as random variables. The fuzzy interval regression analysis model calculates the probability of being in these ranges by determining the lower and upper limit values (confidence interval) of the estimated dependent variable values [5]. Briefly, while the classical regression analysis method is based on probability theory, the fuzzy interval regression analysis approach is based on the possibilistic and fuzzy set theory [6].

One can face some problems in creating a predictive model where the dependent variable is expressed as continuous, but uncertain (i.e. a variable can have categories, classifications or segmentations as low, medium and high) [7]. The fuzzy approach can provide an alternative way to the classical regression analysis. The aim of this study is to calculate the fuzzy coefficient values and their deviations in the interval regression model and interpret them on obstructive sleep apnea syndrome (OSAS). It has been tried to show that more reliable and consistent results can be calculated by applying the interval regression on a small dataset consisting of some measurements for patients with sleep apnea.

Materials and Methods

Data collection

The patients admitted to the Pulmonary Diseases Department of SDU Training and Research Hospital with OSAS (Obstructive Sleep Apnea Syndrom) complaint was the material of the study. The ethical approval was obtained from Süleyman Demirel University, Faculty of Medicine Clinical Research Ethics Committee (01/08/2019 dated and 30856962-813.99-E 234 issue). A total of 514 patients admitted with a suspicion of OSAS between the years of 2005 and 2016 were referred to the sleep clinic by examining with a polysomnography (PSG) device. The patients had snoring, drowsiness during the day, sleep apnea and OSAS symptoms. The patients with a sleep disorders were excluded from the study due to psychiatric reasons, sleep medication use, muscle relaxant use, patients under 18 years of age and other sleep disorder complaints.

Power analysis and sample size

The power analysis was performed by GPower 9.1.2 (Universitaet Kiel, Germany). The test family was selected as F statistics, and the method as "Linear multiple regression: Fixed model, R^2 deviation from zero". The total sample size was determined as 35 with 0.5 effect size, 5% type-I error and 0.85 power for five predictors. Since there are two types of OSAS between the patients as positional and non-positional, then 35 patients for each group were selected randomly. This amount of sample size is already sufficient for interval regression. Because the applications of fuzzy approaches on fewer data sets yielded much better results. Five independent variables: Body Mass Index (BMI = X_1), Rapid Eye Movement (REM = X_2), Oxygen Desaturation Index (ODI = X_3), Arousal (X_4) and Apnea-Hypopnea duration (AH = X_5) were determined as predictors on the dependent variable Apnea-Hypopnea Index (AHI).

Statistical analyses

The classical OLS regression and fuzzy interval regression models were established, and the coefficients were calculated. When the assumptions of the OLS couldn't met or the OLS model is not sufficient to determine the coefficients, a fuzzy interval regression model can be assumed to overcome the modelling problems. This type of model can be used in small samples to minimize the uncertainities of the datasets to determine more accurate results [7,8].

Moreover, mean squared error (MSE), the coefficient of determination (R^2), residual analyses and the problems with multiple linear regression were used to decide the goodness-of fit and the performance of the models. The analyses of the study were performed by Microsoft Office EXCEL 2016 (R), LINGO 16.0 ve TURCOSA Analytic Solutions Version 1.0 softwares.

Interval Regression Analysis Model

Interval regression analysis is one of the fuzzy regression analysis approach and the simplest form of probabilistic regression analysis method [8]. The main feature of the approach is the calculation of fuzzy linear regression coefficients in the range number type. Because of the coefficient values calculated in the range number type, all predicted fuzzy values of the dependent $\tilde{Y}_i$ variable are calculated so that they obtain the most appropriate values in a range with the maximum dispersion level. In other words, the linear regression analysis model, created when the fuzzy tolerance level is h = 0.0, is called the interval regression analysis model [9-12].

In order to apply the interval regression analysis approach to the data sets consisting of exact values, the following systematic path should be followed [8,13];

i. Exact data sets (y_i, x_i), are arranged as i = 1,...,n,

ii. To calculate the values of the coefficient $\tilde{A}_i$, i = 1,2,...,n, of the range number type that minimizes the objective function J, the data is converted to the linear programming problem. For the objective function "J", solving linear programming problem:

$$\min_{a_c, a_s} J = \min_{c,s} \left[s_0 + \sum_{j=1}^{n} a_s^t \left| X_{ij} \right| \right] \tag{1}$$

Topuz and Karaibrahimoğlu. Int J Clin Biostat Biom 2021, 7:037

• Page 2 of 12 •

$$\min_{c,s} J = c_1, c_2, \ldots, c_n, \ c_j \geq 0, \forall i; \ i = 1,2,\ldots,m$$

$$\min_{c,s} J = s_1, s_2, \ldots, s_n, \ s_j \geq 0, \forall i; \ j = 0,1,2,\ldots,n$$

$$c_0 + \sum_{j=0}^{n} c_j X_{ij} + (1 - h)\left[c_0 + \sum_{j=0}^{n} c_j |X_{ij}|\right] \geq \tilde{Y}_i + (1 - h)\tilde{Y}_S \ \forall i; \ i = 1,2,..,n \tag{2}$$

$$c_0 + \sum_{j=0}^{n} c_j X_{ij} - (1 - h)\left[c_0 + \sum_{j=0}^{n} c_j |X_{ij}|\right] \leq \tilde{Y}_i - (1 - h)\tilde{Y}_S \ \forall i; \ i = 1,2,..,n \tag{3}$$

is calculated as above [14,15].

Equation (1) $a_{si} \geq 0$, $a_{ci} \geq 0$, i = 0,1,2,…,n corresponds to the value. The scattering index of the interval regression analysis model is applied to the fuzzy data under the lower limit and upper limit constraints in equation (1,2,3) to minimize total fuzziness. In order to limit the fuzziness of the coefficient values of the model between certain values, the objective function and constraint values must be analyzed together. A reliable solution to the problem is only possible by calculating the coefficient values of the range number type. The lower limit and upper limit constraints;

$$\min_{a_c, a_s} J1 = \begin{bmatrix} c0+X_{i1}*c1+X_{i2}*c2+X_{i3}*c3+X_{i4}*c4+X_{i5}*c5-s0-X_{i1}*s1-X_{i2}*s2-X_{i3}*s3-X_{i4}*s4-X_{i5}*s5<=Y_1-5 \\ c0+X_{i1}*c1+X_{i2}*c2+X_{i3}*c3+X_{i4}*c4+X_{i5}*c5-s0+X_{i1}*s1+X_{i2}*s2+X_{i3}*s3+X_{i4}*s4+X_{i5}*s5>=Y_1+5 \end{bmatrix}$$

$$\ldots\ldots$$

$$\min_{a_c, a_s} Jn = \begin{bmatrix} c0+X_{i35}*c1+X_{i35}*c2+X_{i35}*c3+X_{i35}*c4+X_i*c5-s0-X_i*s1-X_i*s2-X_i*s3-X_i*s4-X_{i35}*s5<=Y_{35}-5 \\ c0+X_{i35}*c1+X_{i35}*c2+X_{i35}*c3+X_{i35}*c4+X_i*c5-s0+X_i*s1+X_i*s2+X_i*s3+X_i*s4+X_{i35}*s5>=Y_{35}+5 \end{bmatrix}$$

form. In the interval number property calculated under certain constraints, coefficient $\tilde{A}_i$, values are displayed as $\tilde{A}_i = (a_{ci}, a_{si})$.

Where a_{ci}, $a_{c0} + a_{c1}X_{i1} + \ldots + a_{c(p-1)}X_{i(p-1)}$, which consists of the values of the middle (center) value and a_{si}; $a_{s0} + a_{s1}|X_{i1}| + \ldots + a_{s(p-1)}|X_{i(p-1)}|$. The values in the form of the coefficient values indicate the half-dispersion value. The coefficient values to be calculated are $\tilde{A}_i$,

$$\tilde{A}_i = \left\{a \,\middle|\, a_{ci} - a_{si} \leq a \leq a_{ci} + a_{si}\right\} \tag{4}$$

is expressed by [14]. That is, by adding the interval operations to the fuzzy regression equation, the coefficients are calculated in the interval number type [16].

i.   Interval regression analysis model to be formed using the coefficient values calculated under certain constraints,

$$\tilde{Y}_i = \tilde{A}_0 + + \tilde{A}_1 X_{i1} + \tilde{A}_2 X_{i2} + \tilde{A}_3 X_{i3} + \cdots + \tilde{A}_{p-1} X_{i(p-1)} = \tilde{A}_i X_i \tag{5}$$

$$\tilde{Y}_i = \{a_{c0}, a_{s0}\} + \{a_{c1}, a_{s1}\} X_{i1} + \{a_{c2}, a_{s2}\} X_{i2} + \cdots + \{a_{c(p-1)}, a_{s(p-1)}\} X_{i(p-1)} \tag{5.a}$$

$$\tilde{Y}_i = \{a_{c0}, a_{s0}\} + \{a_{c1}X_{i1}, a_{s1}|X_{i1}|\} + \cdots + \{a_{c(p-1)}X_{i(p-1)}, a_{s(p-1)}|X_{i(p-1)}|\} \tag{5.b}$$

$$\tilde{Y}_i = \{a_{c0} + a_{c1}X_{i1}, a_{s0} + a_{s1}|X_{i1}|\} + \cdots + \{a_{c(p-1)}X_{i(p-1)}, a_{s(p-1)}|X_{i(p-1)}|\} \tag{5.c}$$

were obtained by [14].

Fuzzy dependent variable $\tilde{Y}_i$ values of the calculated interval number type as a result of multiplying X_{ij} independent (input) variables and model coefficients of interval number type $\tilde{A}_i$, which can represent the data of the investigated properties,

$$\tilde{Y}_i = \left(a_c^t x_j, \ a_s^t |x_j|\right), \tag{6}$$

is expressed as. Here; $a_c = \left(a_{c1}, a_{c2}, a_{c3}, \ldots, a_{cn}\right)^t$, $a_s = \left(a_{s1}, a_{s2}, a_{s3}, \ldots, a_{sn}\right)^t$,

$$|X_i| = \left(|X_1|, |X_2|, \ldots, |X_{ij}|\right)^t$$

Membership function for the dependent variable (Y_i) expressed as equation (6);

$$\mu_{\tilde{Y}}(Y) = \begin{cases} \left\{\tilde{A}_i \,\middle|\, Y = \max_f (\tilde{A}, x)\right\} \min_f \mu_{\tilde{A}_i}(x_i), & \left\{\tilde{A}_i \,\middle|\, Y = f(\tilde{A}, x)\right\} \neq \varnothing \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

is calculated as above. For all calculated fuzzy output $\tilde{Y}_i$ values, the lower and upper limit values generated,

$$a_c^t X_j - a_s^t \left| X_j \right| \leq \tilde{Y}_i \leq a_c^t X_j + a_s^t \left| X_j \right|, j = 1,2,\ldots,n \tag{8}$$

are included in the forecast range [17,18].

According to the extension principle, the $\tilde{Y}_i$ value of each observed dependent variable is estimated as a symmetrical triangular fuzzy number $\tilde{Y}_i = (\tilde{Y}_c, \tilde{Y}_s)$. The limitations of the fuzzy linear regression analysis approach are made at any level h_i determined by the researcher [18-20]. Observation values of the dependent Y_i variable examined to calculate valid and reliable results with fuzzy interval regression analysis approach, the interval boundaries of $\tilde{Y}_i$ dependent variable values estimated as fuzzy interval numbers $Y_i \in \left| \tilde{Y}_i \right|_h$ condition. Under the constraints in equation (1), these values are calculated to predict the dependent variable $\tilde{Y}_i$ values of the lower interval limits $Y_i \in \left| \tilde{Y}_i \right|_h$ I to be within the limits satisfying the condition;

$$\{a_{c0} - a_{s0}\} + \{a_{c1} - a_{s1}\} X_{ij} \leq \tilde{Y}_i^L \forall i; i = 1,2,\ldots,n \tag{9}$$

$$\{a_{c0} + a_{s0}\} + \{a_{c1} + a_{s1}\} X_{ij} \geq \tilde{Y}_i^U \forall i; i = 1,2,\ldots,n \tag{10}$$

The sum of the regression squares (SSR) is the measure of the variation between lower and upper boundaries of $\tilde{Y}_i^{h=1}$. The sum of squared error (SSE) is the measure of the variation of the differences that would occur if the central $\tilde{Y}_i^{h=1}$ value of the dependent variable is used to predict the fuzzy dependent $\tilde{Y}_i$ variable [19,21]. For positional and non-positional OSAS patients, the analyzing constraints and evaluating the models were performed by LINGO 16.0 software.

Results

The classical OLS regression models were established for two OSAS types partitioning the training (71.5%) and test (%28.5) sets by "enter" method, that is, all independent variables were included into the models. In the model of positional OSAS (PPs), only ODI variable was found to be significant (Beta = 0.85; p = 0.001), so as in the model of non-positional OSAS (non-PPs) (Beta = 0.82; p = 0.027). The coeffficients of determination were very low and unacceptable in both models ($R^2 = 0.781$ and $R^2 = 0.889$). When the forward selection method was applied to the models, nothing was changed and R^2 values were a little bit greater. As seen in Table 1, the coefficients of determination were too low and MSE values of the models were very high. Therefore, the models could be considered as unfitted. Moreover, the multiple linear regression assumptions were checked, and the residual analyses were performed. There were no autocorrelation and heteroscedasticity problems in both of the models. The Durbin-Watson statistics were found as 1.89 (positional) and 2.02 (non-positional) for the models. The standardized predicted vs. residual scatter plots indicated no heteroscedasticity (Figure 1 and Figure 2). Although the highest VIF values were low ($VIF_{positional} = 1.421$, $VIF_{non-positional} = 1.713$) for the models, the condition index values exceeded the threshold as 98.39 and 113.86. This showed that there is a slightly higher multicollinearity in the models. Thus, this explained why R^2 values were high although most of the independent variables were not significant. In residual analysis, dFits, dfBeta and Cook's distance measures showed that there was only one outlier observation in each model.

Table 1: The coefficient of the classical OLS regression models (n = 35).

	Coefficients for positional OSAS			Coefficients for non-positional OSAS		
	Training set			**Training set**		
Variables	β	Standard Error	p	β	Standard Error	p
Constant	123.96	99.08	0.226	-216.80	226.32	0.392
BMI (X_1)	-4.10	3.29	0.229	9.01	9.24	0.385
REM (X_2)	0.14	0.21	0.520	-1.33	1.32	0.370
ODI (X_3)	0.85	0.11	0.001*	0.82	0.71	0.027*
AROUSAL (X_4)	-0.13	0.08	0.128	0.01	0.22	0.939
AH (X_5)	0.05	0.02	0.082	-0.04	0.33	0.912
Model (Enter)		$R^2 = 0.781$	< 0.001*		$R^2 = 0.889$	0.001*
		MSE = 36.27			MSE = 34.53	
	Test set			**Test set**		
Constant	107.89	308.65	0.744	6.11	62.60	0.923
BMI (X_1)	-3.17	10.41	0.776	0.13	2.32	0.956
REM (X_2)	-0.31	0.41	0.494	-0.41	0.26	0.133

ODI (X_3)	0.54	0.28	0.122	0.84	0.11	< 0.001[*]
AROUSAL (X_4)	-0.24	0.28	0.447	0.01	0.09	0.985
AH (X_5)	0.06	0.07	0.412	0.02	0.03	0.629
Model (Enter)		R^2 = 0.679	0.123		R^2 = 0.719	< 0.001[*]
		MSE = 28.69			MSE = 6.23	

[*]Significant at 0.05 level; AHI: Apnea-Hypopnea Index (event/h); BMI: Body Mass Index (kg/m^2); REM: Rapid Eye Movement Sleep (%); ODI: Oxygen Desaturation Index (event/h); AROUSAL: Stimulating in sleep (event/h); AH: Apnea-Hypopnea duration (h); R^2: Coefficient of determination; MSE: Mean Squared Error; RMS: Root Mean Squared Error.
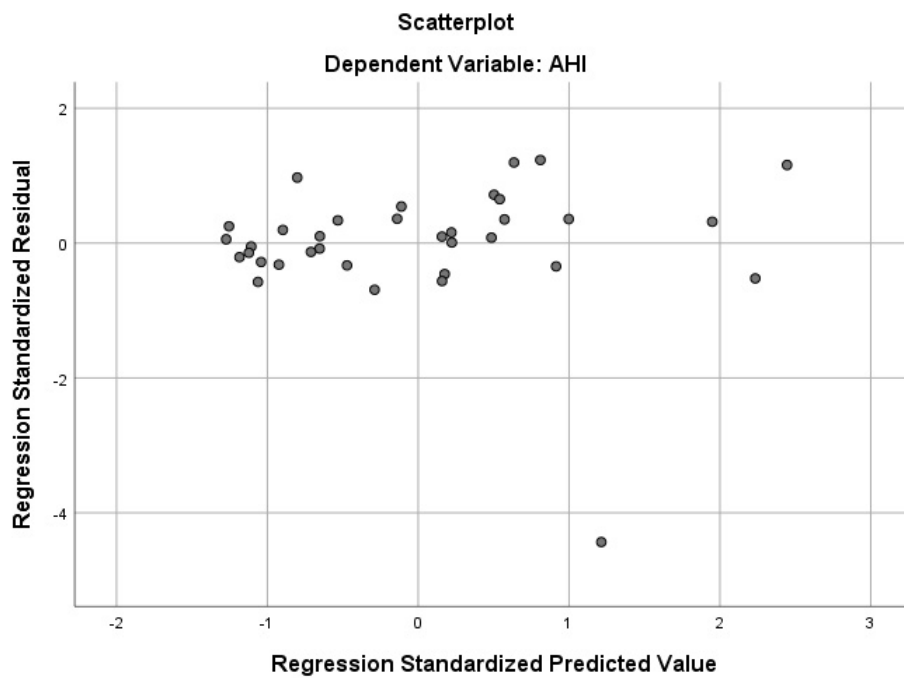

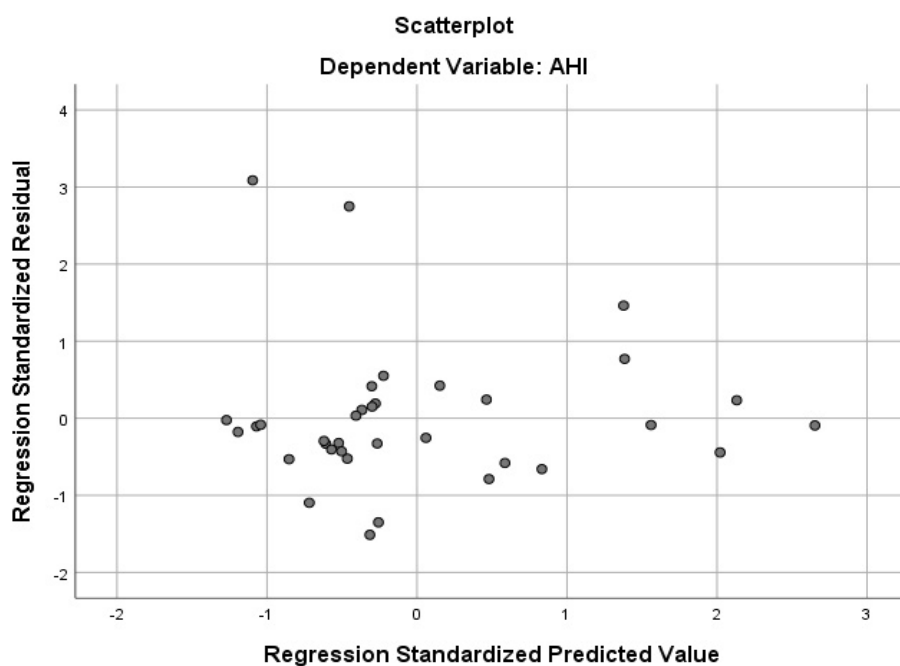**Figure 1:** Scatter plot of predicted vs. residual of positional OSAS model.

Figure 2: Scatter plot of predicted vs. residual of non-positional OSAS model.

Table 2: The center and dispersion values of coefficients calculated at h = 0.0 fuzziness tolerance level for models of positional OSAS patients.

Dependent Variables	$\tilde{A}_j = (a_{cj}; a_{sj})$ h = 0.0 level						
		Test set (n = 10)		Training set (n = 25)		n = 35	
	Coefficients	Center value (a_{cj})	Dispersion (a_{sj})	Center value (a_{cj})	Dispersion (a_{sj})	Center value (a_{cj})	Dispersion (a_{sj})
Constant	$\tilde{A}_0$	109.72	0.00	-24.25	0.58	1.5	1.47
BMI (X_1)	$\tilde{A}_1$	-3.44	0.11	0.72	0.09	-0.08	0.11
REM (X_2)	$\tilde{A}_2$	-0.28	0.13	0.16	0.19	0.01	0.08
ODI (X_3)	$\tilde{A}_3$	0.81	0.02	1.07	0.05	1.07	0.04
AROUSAL (X_4)	$\tilde{A}_4$	-0.23	0.23	0.04	0.01	0.12	0.11
AH (X_5)	$\tilde{A}_5$	0.53	0.00	0.02	0.01	0.02	0.01

Table 3: Estimated values of AHI (Apnea-Hypopnea Index) for non-positional OSAS patients.

No	Observed AHI values (event/h) (Y_i)	Estimated AHI ($\tilde{Y}_i$) statistics on values			
		$\tilde{Y}_c$	$\tilde{Y}_s$	Lower fuzziness limit values	Upper fuzziness limit values
1	5.0	5.584	2.722	2.862	8.307
2	70.80	67.917	28.301	39.616	96.218
3	35.8	35.083	12.830	22.253	47.914
4	7.9	9.968	3.036	6.931	13.005
.	…	…	...	...	...
33	63.9	66.514	25.943	40.570	92.457
34	20.2	22.584	9.418	13.166	32.003
35	30.1	38.976	13.048	25.928	52.025
Mean	26.57	29.27		17.999	39.833
SEM	2.99	3.08			
Min-Max	5.0-70.8	1.58-68.41	2.08-28.30	-0.49-42.98	3.67-96.21
Mean difference			t statistics	p	
-0.37			-0.293	0.772	

Prediction of AHI values by fuzzy interval regression analysis

Using these values in the interval regression analysis method, fuzzy statistical values such as center values, dispersion, lower and upper limit values and MSE values were calculated.

i) As a result of following the systematic stages mentioned in the method section, the data was arranged.

ii) Equation (11) for the model to be at the minimum fuzziness level coefficient and dispersion values using the objective function in (1) for 35 Positional OSAS under 70 OSAS patients (both groups) with h = 0.0 fuzziness tolerance level constraint using;

$$\min_{a_c, a_s} J1 = \begin{bmatrix} c0+29.0*c1+11.60*c2+5.45*c3+1.10*c4+14.2*c5-s0-29.0*s1-11.63*s2-5.45*s3-1.10*s4-14.2*s5<=5.0-5 \\ c0+29.0*c1+11.60*c2+5.45*c3+1.10*c4+14.2*c5-s0+29.0*s1+11.63*s2+5.45*s3+1.10*s4+14.2*s5>=5.0+5 \end{bmatrix}$$

…..

$$\min_{a_c, a_s} J35 = \begin{bmatrix} c0+29.4*c1+9.50*c2+17.86*c3+10.0*c4+48.8*c5-s0-29.4*s1-9.50*s2-17.86*s3-10.0*s4-48.8*s5<=17.90-5 \\ c0+29.4*c1+9.50*c2+17.86*c3+10.0*c4+48.8*c5-s0+29.4*s1+9.50*s2+17.86*s3+10.0*s4+48.8*s5>=17.90+5 \end{bmatrix}$$

conditions for coefficient values were created (Table 2).

iii) Equation created using coefficient values in Table 2, calculated at h = 0.0 fuzziness tolerance, interval regression analysis model in (11);

$$\tilde{Y}_i = \tilde{A}_0 + \tilde{A}_1 X_{i1} + \tilde{A}_2 X_{i2} + \tilde{A}_3 X_{i3} + \tilde{A}_4 X_{i(4)} + \tilde{A}_5 X_{i(5)} \tag{11}$$

$\tilde{Y}_i$ = {1.5646; 1.4724} + {-0.0799; 0.1184}$X_{i(1)}$ + {0.0091; 0.0891}$X_{i(2)}$ + {1.0768; 0.0419}$X_{i(3)}$ + {0.1276; 0.1153}$X_{i(4)}$ + {0.0109; 0.0121}$X_{i(5)}$

was yielded.

iv) The value of the objective function Z(x) representing the fuzziness level of equation (11), as a result of the application of the dispersion values calculated in Table 3 as in equation (12),

$$Z = 2 \times \left[35\, a_0^S + a_1^S \sum_{i=1}^{35} x_{1i} + a_2^S \sum_{i=1}^{35} x_{2i} + a_3^S \sum_{i=1}^{35} x_{3i} + a_4^S \sum_{i=1}^{35} x_{4i} + a_5^S \sum_{i=1}^{35} x_{5i} \right] \tag{12}$$

$$Z = 2 \times \left[35 \times a_0^S + 1035.5x\, a_1^S + 476.2x\, a_2^S + 977.59x\, a_3^S + 830.2x\, a_4^S + 2769.3x\, a_5^S \right]$$

$$Z = 2 \times \left[35 \times 1.4724 + 1035.5 \times 0.1184 + 476.2 \times 0.0891 + 977.59 \times 0.0419 + 830.2 \times 0.1153 + 2769.3 \times 0.0121 \right] = 773.9$$

V) Using the equation (11) generated at the fuzziness level of Z = 773.9, the estimated $(\tilde{Y}_c)$ AHI (Apnea-Hypopnea Index) values for the 35 Positional OSAS (PPs) patients and the lower and upper fuzziness limit values were calculated by MS-Office EXCEL (Table 4). There was no significant difference between the observed and predicted AHI values (p = 0.672) according to the calculated values at h = 0.0 fuzzy level in Table 4. MSE values for training, test and whole sets were found as 13.62, 11.08 and 28.16 respectively.

Moreover, the observed and predicted AHI (event/h) values for both model was compared by paired sample t-test. There was no significant difference between the predicted and observed AHI values for both positional and non-positional OSAS patients.

The observed $(\tilde{Y}_i)$ and the estimated $(\tilde{Y}_c)$ AHI values were shown in Figure 3. It can be seen that the observed and estimated AHI values for PPs patients exactly coincided for 1st, 8th, 21st, 22nd, 23rd, 29th, 30th, 32nd, and 34th observations.

It can also be seen that there is a significant correlation between the observed and the estimated values of AHI with r = 0.994 in Figure 4.

Table 4: Estimated values of AHI (Apnea-Hypopnea Index) for positional OSAS patients.

| No	Observed AHI values (event/h) (Yᵢ)	Estimated AHI $(\tilde{Y}_i)$ statistics on values			
		$\tilde{Y}_c$	$\tilde{Y}_s$	Lower fuzziness limit values	Upper fuzziness limit values
1	5.0	5.515	6.468	0.952	11.983
2	29.5	31.840	9.041	22.799	40.882
3	16.0	17.267	9.501	7.766	26.768
4	52.5	51.027	11.114	39.913	62.142
…	…	…	...	...	...
33	32.7	35.168	11.769	23.399	46.938
34	9.3	9.731	7.476	2.254	17.207
35	17.9	20.343	8.295	12.047	28.638
Mean	29.07	32.89		22.203	43.694
SEM	3.39	3.80			
Min-Max	5.0-83.4	5.51-95.22	6.46-19.05	-1.87-78.27	11.98-112.16
Mean difference			t statistics	p	
-3.12			-0.42	0.672	
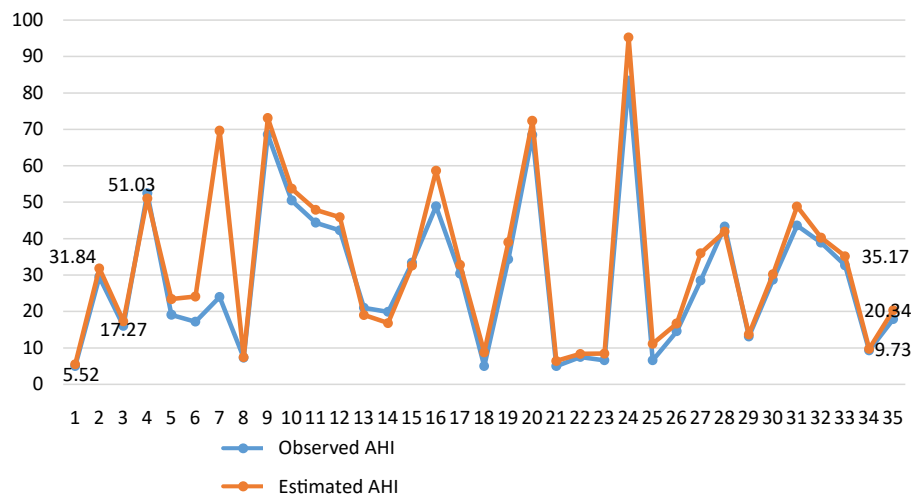
Figure 3: Line graphs of the observed and estimated $(\tilde{Y}_C)$ AHI values for positional OSAS patients.
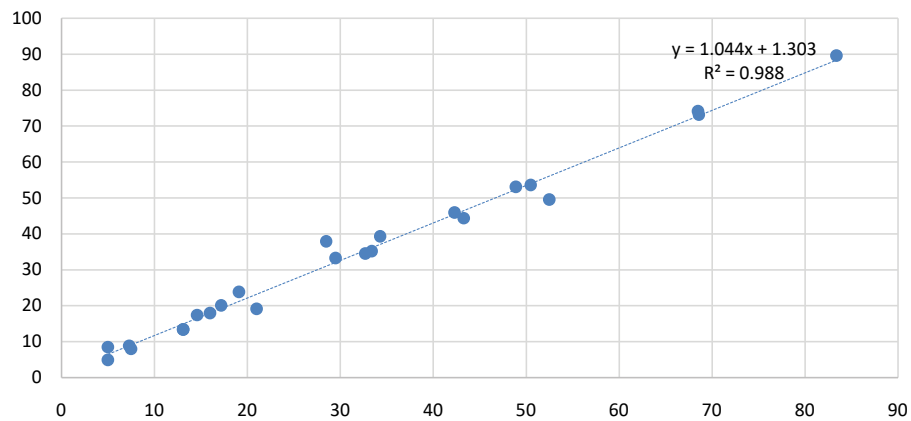

Figure 4: The correlation between the observed and estimated mean $(\tilde{Y}_C)$ AHI values for PPs patients.

i) Similarly, the data for non-positional OSAS patients were analyzed

ii) Using the objective function in equation (1), (13) was obtained for 35 non-positional OSAS patients with h = 0.0 tolerance level and under the constraint 70 (35 observations*2) patients;

$$\min_{a_c, a_s} J1 = \begin{bmatrix} c0+25.8*c1+21.2*c2+4.85*c3+34.10*c4+30.0*c5-s0-25.8*s1-21.2*s2-4.85*s3-34.10*s4-30.0*s5<=5.0-5 \\ c0+25.8*c1+21.2*c2+4.85*c3+34.10*c4+30.0*c5+s0+25.8*s1+21.2*s2+4.85*s3+34.10*s4+30.0*s5>=5.0+5 \end{bmatrix}$$

$$\cdots\cdots$$

$$\min_{a_c, a_s} J35 = \begin{bmatrix} c0+26.8*c1+3.3*c2+29.6*c3+27.5*c4+96.9*c5-s0-26.8*s1-3.3*s2-29.6*s3-27.5*s4-96.9*s5<=30.1-5 \\ c0+26.8*c1+3.3*c2+29.6*c3+27.5*c4+96.9*c5+s0+26.8*s1+3.3*s2+29.6*s3+27.5*s4+96.9*s5>=30.1+5 \end{bmatrix}$$

conditions for coefficient values were created. For non-positional apnea patients, the constraints created at h = 0.0 fuzziness tolerance level were analyzed and coefficient values of equality (3.4) $\tilde{A}_j$, $j = 0, \ldots, 5$ were calculated as in Table 5.

iii) Equation created using coefficient values in Table 5, calculated at h = 0.0 fuzziness tolerance, interval regression analysis model in 13;

$$\tilde{Y}_i = \tilde{A}_0 + \tilde{A}_1 X_{i1} + \tilde{A}_2 X_{i2} + \tilde{A}_3 X_{i3} + \tilde{A}_4 X_{i(4)} + \tilde{A}_5 X_{i(5)} \tag{13}$$

$$\tilde{Y}_i = \{-142.7171;\ 0.0000\} + \{5.5233;\ 0.0000\} X_{i1} + \{-0.2376;\ 0.3053\} X_{i2} + \{0.6667;\ 0.0000\} X_{i3} +$$
$$\{0.1191;\ 0.0414\} X_{i(4)} + \{0.1181;\ 0.1306\} X_{i(5)}$$

iv) The value of the objective function Z (x) representing the fuzziness level of equation 13, applying the dispersion values calculated in Table 4 as in Equation (14),

Table 5: The center and dispersion values of coefficients calculated at h = 0.0 fuzziness tolerance level for models of non-positional OSAS patients.

Variables	$\tilde{A}_j = (a_{cj}; a_{sj})$ h = 0.0 level						
	Test set (n = 10)			Training set (n = 25)		n = 35	
	Coefficients	Center value (a_{cj})	Dispersion (a_{sj})	Center value (a_{cj})	Dispersion (a_{sj})	Center value (a_{cj})	Dispersion (a_{sj})
Constant	$\tilde{A}_0$	-305.17	0.0	-138.51	0.0	-142.71	0.0
BMI (X_1)	$\tilde{A}_1$	12.76	0.0	5.34	0.0	5.52	0.00
REM (X_2)	$\tilde{A}_2$	-1.99	0.39	-0.24	0.33	-0.24	0.30
ODI (X_3)	$\tilde{A}_3$	1.45	0.0	0.65	0.0	0.6	0.00
AROUSAL (X_4)	$\tilde{A}_4$	0.16	0.0	0.14	0.01	0.11	0.04
AH (X_5)	$\tilde{A}_5$	-0.26	0.04	0.12	0.11	0.11	0.13
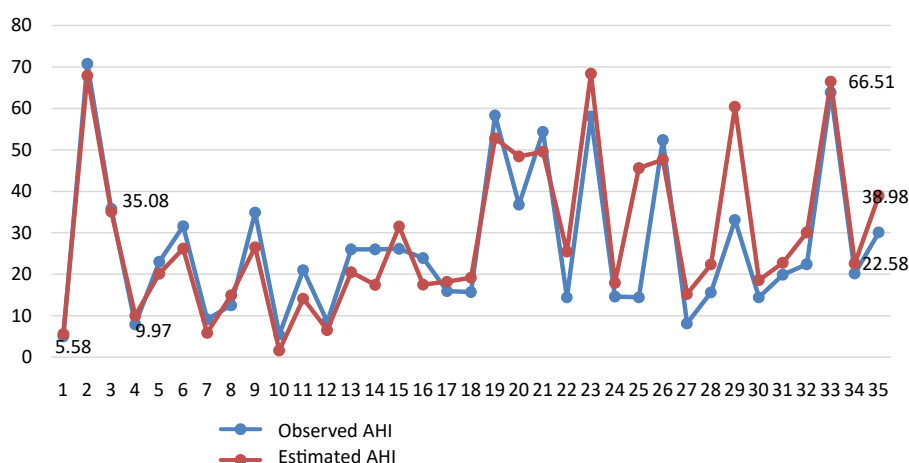
Figure 5: Line graphs of the observed and estimated $(\tilde{Y}_C)$ AHI values for PPs patients.

$$Z = 2x \left[35\, a_0^S + a_1^S \sum_{i=1}^{35} x_{1i} + a_2^S \sum_{i=1}^{35} x_{2i} + a_3^S \sum_{i=1}^{35} x_{3i} + a_4^S \sum_{i=1}^{35} x_{4i} + a_5^S \sum_{i=1}^{35} x_{5i} \right] \quad (14)$$

$$Z = 2\, x \left[35\, x\, a_0^S + 917.1\, x\, a_1^S + 401.9\, x\, a_2^S + 835.02\, x\, a_3^S + 989.0\, x\, a_4^S + 3071.5\, x\, a_5^S \right]$$

$$Z = 2\, x \left[35\, x\, 0.000 + 917.1\, x\, 0.000 + 401.9\, x\, 0.3053 + 835.02\, x\, 0.0000 + 989.0\, x\, 0.0414 + 3071.5\, x\, 0.1306 \right] = 1129.54$$

Estimated mean $(\tilde{Y}_C)$ AHI values, and lower and upper fuzziness limit values for 35 non-positional apnea patients using the equation (13) generated at fuzziness level at Z = 1129.54 were calculated and presented in Table 3. There was no significant difference between the observed and predicted AHI values (p = 0.772) according to the calculated values at h = 0.0 fuzzy level in Table 3. MSE values for training, test and whole sets were found as 25.62, 5.94 and 23.70 respectively.

The mean $(\tilde{Y}_i)$ values and the estimated mean $(\tilde{Y}_C)$ AHI values coincided at 1st, 3rd, 12nd and 33rd observations (Figure 5)

There was a significant correlation (r = 0.957) between observed and estimated mean $(\tilde{Y}_C)$ AHI values for non-positional apnea patients (Figure 6).
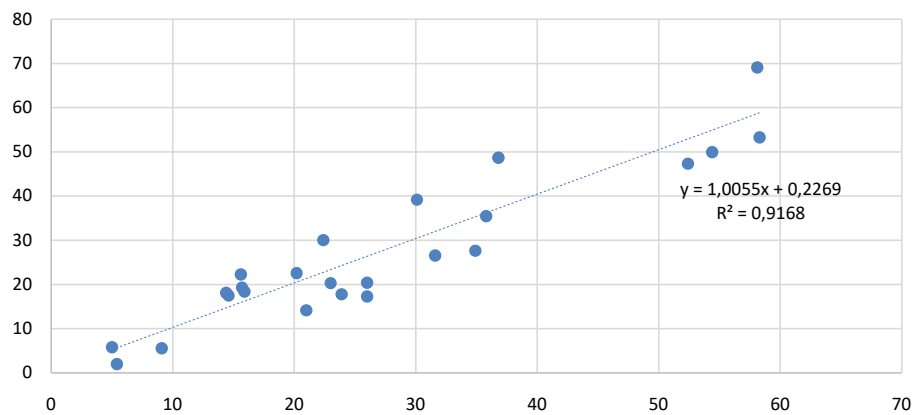
Topuz and Karaibrahimoğlu. Int J Clin Biostat Biom 2021, 7:037

● Page 9 of 12 ●

Figure 6: The correlation between the observed and estimated mean $(\tilde{Y}_C)$ AHI values for non-PPs patients.

On the evaluation of the interval regression models, it can be said that the model for PPs was a better and more reliable as prediction. Because, there were three strongly significant (REM, ODI, AH-duration) and two moderate significant (BMI & AROUSAL) independent variables. The dispersion values of the strongly significant variables were close to zero. This means that the deviation is so small in the fuzziness interval. Therefore, the total fuzziness level value was lower and found as Z(x) = 773.9. Among the significant variables, ODI had a great contribution to explain AHI with a center value of 1.07. For example, AH-duration was significant on AHI but had a lower effect. For the model of non-PPs, the most significant variable was BMI, and had an effect of 5.52 on AHI. ODI was another significant variable on AHI with 0.0 dispersion. Only REM and AH-duration variables were moderate significant, and had higher dispersion values. According to the dispersion values, the objective function Z(x) = 1129.54 as a fuzziness level. Therefore, it can be seen that more reliable prediction values were obtained by calculating the average coefficients with minimum deviation at maximum uncertainty (fuzziness) level.

Discussion

The application of interval regression in the field of health is not widespread. However, in some areas one can see many applications, especially in support vector regression applications. In a study, the support vector regression (SVR) approach with ε-insensitive loss function was proposed to solve a linear constrained quadratic programming problem in a training dataset [22]. Similarly, Hong and Hwang [23] studied interval regression using quadratic loss support vector machine, and it was claimed that the suggested method is attractive for nonlinear interval data. In another study, SVR with interval regression strategy was proposed to estimate the interval bounds for both linear and non-linear regressions in a simulated crisp dataset [24]. Hladik and Cerny [25] adapted tolerance analysis for interval regression in their study, and solved the

problem of outliers in a crisp data. They, also, claimed that the method was more flexible to control the relative widths of interval parameters and simple to compute the results, even for large regression models. In a real house price dataset concerning the 32 cities in Korea, interval regression based on quantile techniques was proposed for upper and lower approximation modeling [26].

In some studies, fuzzy regression was extended into an approach in which the dependent and independent variables are both fuzzy [27]. In a recent study which is a little bit similar to this study, the authors dealt with the interval regression with endogenous independent variables. There were two applications: First one was a simulation study, second one was on a real survey data describing the behaviour of immigrants' remittance. Age, gender, education level, household income, presence of close relatives and intention to return to the home country were independent variables whereas the three remittance intervals (< 1000, 1001-5000 and > 5001 AUS $) was the dependent variable. It was shown that the method was effective for endogeneity of income and consumption in the context of remittance' behaviour [5].

In this study, valid and reliable results were calculated by interval regression model to determine the estimated values of AHI for patients with obstructive sleep apnea. OSAS is a disorder caused by respiratory obstruction during sleep. It has been reported to be the cause of many complications from cardiovascular diseases to obesity. In the advanced stage, it results in cerebrovascular events and sudden deaths. There are generally two phenotypes, positional and non-positional. Positional OSAS is a condition in which apnea occurs during sleep on their backs. OSAS, which is not related to the position, is a clinical condition that is mostly associated with REM sleep or characterized by arousals as a result of obstruction of the upper airway [28]. Polysomnography (PSG) is the gold standard diagnostic method for Obstructive Sleep Apnea Syndrome.PAP

Topuz and Karaibrahimoğlu. Int J Clin Biostat Biom 2021, 7:037

• Page 10 of 12 •

(Positive Airway Pressure) is the gold standard method of treatment. Although it is recommended mainly in patients with moderate and severe OSAS (AHI > 15), PAP treatment is recommended if comorbid diseases are present, even in mild patients. RDI (Respiratory Disturbance Index) is the most important factor showing position dependence in OSAS patients. Obesity, age or concomitant respiratory or cardiovascular diseases are also among the risk factors. However, a clear distinction between positional (PPs) and non-positional OSAS (non-PPs) patients has not yet been established. PPs patients can turn into non-PPs patients as violence increases because non-PPs patients have a more disturbed sleep structure [29,30].

Worsening of breathing while sleeping in the supine position is more common in male patients. Although one of the reasons is obesity, the increase in respiratory events is also interestingly observed in normal-weight patients. This is probably due to the fact that the tongue falls back due to gravity in the pharyngeal region [31]. In general, better sleep quality and less arousal are observed in the PPs group compared to the non-PPs group. However, despite higher RDI and less frequent oxygen saturation, severe OSA is seen in the supine position [32].

In this study, classical OLS regression and fuzzy interval regression models were compared. Prediction models for AHI values were established with the measurements obtained from polysomnography device for positional and non-positional types for OSAS patients. Since there was no interval regression analysis study conducted in the field of health, the results obtained could not be compared. However, in the study, the superiority of the fuzzy interval regression method was tried to be shown by comparing two different methods. In this case, the OLS regression analysis cannot give accurate results for such a dataset. The OLS models cannot be considered as sufficient, and the independent variables except for ODI are not significant and MSE values are higher than the interval regression models. Therefore, applying the fuzzy regression analysis approach allows experts to make more valid and reliable results. Especially, the interval regression can give us a more reliable and consistent model if there is a collinearity problem in a model or outlier observations in a dataset. The only ODI was found to have a significant effect on the AHI of both PPs and non-PPs patients in OLS models. In fuzzy interval regression models, all the affecting variables have a significant contribution to the model. It is observed that the MSE values of the models are quite low. It can be also shown that the estimated AHI values obtained with fuzzy models were substantially equal to the observation values.

There are, of course, some limitations in this study. The first limitation is the sample size of the study. The interval regression method works in smaller datasets better, and we have selected 35 patients for each groups for both OLS and interval regression models. However, there were 514 eligible patients in the dataset. Therefore, it can be determined different coefficients and significance values for different samples. The second limitation is that there is no common comparable measure between the OLS and the interval regression method. It is necessary to calculate some similar measures between the models. This situation causes to make some additional analyses. Moreover, the interval regression is a long process, and if the sample size is big then one should calculate the fuzziness tolerance values as the sample size.

Conclusion

The change in current problems implies the researches to have a more complex structure. This case reveals the necessity of changing the structure of classical regression models with different techniques. Therefore, new alternative approaches have been developed and an interval regression analysis approach has been applied in many areas. It can be explained that if the dependent variable is continuous and uncertain, and independent variables consist of continuous variables measured at all levels, it can be used as a discrimination model in the grouping of observations. In this study, it was shown that the interval regression analysis model achieved in estimating the dependent variable (AHI) in OSAS patients' dataset. Therefore, the factors affecting the PPs and non-PPs can be detected more carefully in diagnosis, and the treatment can be planned more effective for OSAS. It can be concluded that the researchers studying in the healthcare area are able to use the interval regression modeling to get more accurate results under appropriate conditions.

References

1. Bonett DG, Wright TA (2011) Sample size requirements for multiple regression interval estimation. Journal of Organizational Behavior 32: 822-830.

2. Long JS (1997) Regression models for categorical and limited dependent variables. Sage Publications, Thousand Oaks, CA.

3. Peters G (1994) Fuzzy linear regression with fuzzy intervals. Fuzzy Sets Syst 63: 45-55.

4. Hsien-Chung W (2003) Linear regression analysis for fuzzy input and output data using the extension principle. Computers and Mathematics with Applications 45: 1849-1859.

5. Bettin G, Lucchetti R (2012) Interval regression models with endogenous explanatory variables. Empirical Economics 43: 475-498.

6. Ross TJ (2004) Fuzzy logic with engineering applications. John Willey and Sons Inc, New York.

7. Tanaka H, Lee H (1998) Interval regression analysis by quadratic programming approach. IEEE Transactions on Fuzzy Systems 6: 473-481.

8. Ishibuchi H (1992) Fuzzy regression analysis. Japan Journal Fuzzy Theory and Systems 4: 137-148.

9. Moore RE (1966) Interval analysis. Prentice-Hall, Englewood Cliffs, NJ.

10. Moore RE (1979) Methods and applications of interval analysis. SIAM, Philadelphia.

11. Jajuga K (1986) Linear fuzzy regression. Fuzzy Sets and Systems 20: 343-353.

12. Redden DT, Woodall WH (1994) Properties of certain fuzzy linear regression methods. Fuzzy Sets and Systems 64: 361-375.

13. Chang PT, Lee ES (1994) Fuzzy linear regression with spreads unrestricted in sign. Computers and Mathematics with Applications 28: 61-71.

14. Slowinski R (1998) Fuzzy sets in decision analysis, operations research and statistics. The Handbooks of Fuzzy Sets, London.

15. Arnold SF (1990) Mathematical statistics. Prentice-Hall, New Jersey.

16. Chang YHO, Ayyub BM (2001) Fuzzy regression methods - a comparative assessment. Fuzzy Sets and Systems 119: 187-203.

17. Kacprzyk J, Fedrizzi M (1992) Possibilistic regression analysis based on linear programming. Fuzzy Regression Analysis, Physica-Verlag, Heidelberg, 47-60.

18. Topuz D, Şahinler N (2020) Analysis of categorical data obtained for estimation of royal jelly amount with fuzzy range regression model. Turkish Journal of Agricultural and Natural Sciences 7: 471-479.

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

20. Tanaka H, Hayashi I, Watada J (1989) Possibilistic linear regression analysis for fuzzy data. European Journal of Operational Research 40: 389-396.

21. Sawaka M, Yano H (1992) Multiobjective fuzzy linear regression analysis for fuzzy input-output data. Fuzzy Sets and Systems 47: 173-181.

22. Jeng J, Chuang CC, Su SF (2003) Support vector interval regression networks for interval regression analysis. Fuzzy Sets and Systems 138: 283-300.

23. Hong DH, Hwang C (2005) Interval regression analysis using quadratic loss support vector machine. IEEE Transactions on Fuzzy Systems. 13: 229-237.

24. Hwang C, Hong DH, Seok KH (2006) Support vector interval regression machine for crisp input and output data. Fuzzy Sets and Systems 157: 1114-1125.

25. Hladik M, Cerny M (2012) Interval regression by tolerance analysis approach. Fuzzy Sets and Systems 193: 85-107.

26. Lee H, Tanaka H (1999) Upper and lower approximation models in interval regression using regression quantile techniques. European Journal of Operational Research 116: 653-666.

27. Hojati M, Bector CR, Smimou K (2005) A simple method for computation of fuzzy linear regression. European Journal of Operational Research 166: 172-184.

28. Lee SA, Paek JH, Chung YS, Kim WS (2017) Clinical features in patients with positional obstructive sleep apnea according to its subtypes. Sleep Breath 21: 109-117.

29. Oksenberg A, Silverberg DS, Arons E, Radwan H (1997) Positional versus non-positional obstructive sleep apnea patients. Sleep and Breathing 112: 629-639.

30. Karaibrahimoglu A, Ozturk O (2020) Re-evaluation of subtypes of positional OSAS by clustering algorithms. Progress in Nutrition 22: e2020006.

31. Cartwright RD (1984) Effect of sleep position on sleep apnea severity. Sleep 7: 110-114.

32. Teerapraipruk B, Chirakalwasan N, Simon R, Hirunwiwatkul P, Jaimchariyatam N, et al. (2012) Clinical and polysomnographic data of positional sleep apnea and its predictors. Sleep Breath 16: 1167-1172.

Topuz and Karaibrahimoğlu. Int J Clin Biostat Biom 2021, 7:037

• Page 12 of 12 •