ICIC Express Letters
Part B: Applications
Volume **9**, Number **4**, April **2018**

ICIC International ©2018 ISSN 2185-2766
pp. **309–316**

ANALYSIS OF PREFECTURAL CITIZENS' LEISURE TIME USING INTERVAL TYPE REGRESSION MODEL

Yoshiyuki Yabuuchi

Faculty of Economics
Shimonoseki City University
2-1-1 Daigaku-cho, Shimonoseki, Yamaguchi 751-7510, Japan
yabuuchi@shimonoseki-cu.ac.jp

Received October 2017; accepted January 2018

Abstract. *Regression models are typically used for social science field analysis. Additionally, soft computing methods are advantageous when data involve vagueness represented by individuals in a field of social science. However, it is sometimes difficult to illustrate the trend of a target system involving vague data. In this study, interval type regression is improved to manage vagueness in the possibility grade. The improved model analyzes how prefectural citizens spend their free time. The results indicate that economic situations influence the leisure time of individuals. Conversely, health conditions, employment situations, and security do not influence the active spending of leisure time for individuals. The results are discussed in detail.*

Keywords: Fuzzy regression model, Interval type, Possibility grade, Leisure time

1. **Introduction.** Many systems in social fields are constituted of individuals. Hence, vagueness as typified by individuals is mixed, and the relationship between the variables is easily blurred. Generally, various filters are used to deal with this issue. However, there are cases in which filters are unable to sufficiently deal with the same. A model that uses a soft computing method is effective in solving this problem [1-4,9]. Additionally, interval type regressions that use a soft computing method can easily and intuitively understand the possibility of the focal system [13]. An interval type model illustrates the possibility of the focal system by including data. However, the model may be distorted by the shape of the data distribution. Therefore, several results obtained by extant studies are related to the shapes and outliers of an interval type regression [5-8,10,11,14-21].

In an interval type regression, the focus of a study changes to correspond to an outlier and to fit a model to a data distribution. Nevertheless, improvements are continuously required. However, a model that assumes that the possibility grade includes errors was significantly made improved in the conventional model [21]. The improved model eliminates the influence of outliers and illustrates system possibility. The first objective of this paper involves analyzing how prefectural citizens spend their leisure time by using three regression models, namely least squares, interval type, and improved interval type regressions. The second objective involves confirming the usefulness of the improved interval type regression.

This paper is divided into four sections. Section 2 outlines an interval type and an improved interval type regression. Section 3 analyzes ways to spend leisure time according to the prefecture by using three models. Section 4 presents the conclusions of this paper.

2. **Interval Type Regression Model to Manage Vagueness Included in Possibility Grade.** In this paper, an interval type regression is considered with symmetric triangular fuzzy numbers as regression coefficients. In n pairs of $p+1$ variables $(\mathbf{x}_i, y_i)$

$(i = 1, 2, \dots, n)$, $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ denotes an independent variable, and y_i denotes a dependent variable. The regression equation that is obtained is as follows:

$$\mathbf{Y}_i = \mathbf{A}_1 x_{i1} + \mathbf{A}_2 x_{i2} + \cdots + \mathbf{A}_p x_{ip} = (a_1, c_1)x_{i1} + (a_2, c_2)x_{i2} + \cdots + (a_p, c_p)x_{ip}. \tag{1}$$

Here, $x_{i1} = 1$, (a_1, c_1) denotes a constant term that is denoted by the central value a_1 and the width c_1. The center denoted by Y_i^C, width denoted by W_i, and upper and lower limits that correspond to Y_i^U and Y_i^L, respectively, of the interval type regression are as follows:

$$\left.\begin{array}{l} Y_i^C = a_1 x_{i1} + a_2 x_{i2} + \cdots + a_p x_{ip}, \\ W_i = c_1|x_{i1}| + c_2|x_{i2}| + \cdots + c_p|x_{ip}|, \\ \mathbf{Y}_i = \left(Y_i^L, Y_i^C, Y_i^U\right) = \left(Y_i^C - W_i, Y_i^C, Y_i^C + W_i\right). \end{array}\right\} \tag{2}$$

This interval type regression illustrates the possibility of the focal system by including the data, and thus observed values and a model possess an inclusion relation, $y_i \subseteq \mathbf{Y}_i$. Additionally, a decrease in the vagueness of a model, i.e., a decrease in the width, improves the model, and an interval type regression is rewritten as the following linear programming (LP) problem.

$$\left.\begin{array}{ll} \min\limits_{\mathbf{a},\mathbf{c}} & \sum\limits_{i}^{n} W_i \\ \text{s.t.} & y_i \subseteq \mathbf{Y}_i \ (i = 1, 2, \dots, n) \end{array}\right\} \tag{3}$$

The possibility grade $\mu(\mathbf{x}_i, y_i)$ from $\mathbf{Y}_i$ obtained by the regression Equation (1) and sample $(\mathbf{x}_i, y_i)$ is expressed as follows:

$$\mu(\mathbf{x}_i, y_i) = \max\left(0, 1 - \left|y_i - Y_i^C\right| / W_i\right). \tag{4}$$

Thus, the possibility grade is automatically obtained if a model is obtained. Therefore, the shape of the model is easily distorted based on the membership function and data distribution. Hence, it is assumed that vagueness is mixed in possibility grades as well as in independent and dependent variables. The study considers flexibly improving the vagueness included in the possibility grades in the improved regression.

It is assumed that μ_i denotes the possibility grade obtained from the model and samples. The original dependent variable excluding vagueness corresponds to y_i^* when vagueness e_i is mixed in the μ_i. This is illustrated in Figure 1. As shown in Figure 1, y_i^* is farther from the center when e_i corresponds to a positive value. Similarly, it is close to the center if e_i

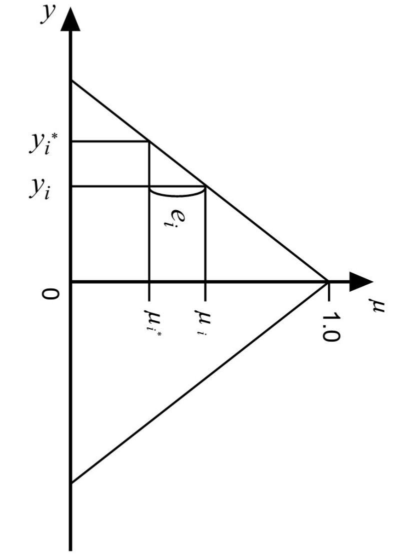


Figure 1. Vagueness included in a possibility grade

is negative. The shape of the model changes if a sample moves from near the boundary of the data distribution to near the center. Hence, the model distortion decreases when the shape of a model is distorted.

The relationship between observation values y_i and y_i^* after removing vagueness corresponds to $y_i^* = y_i + e_i\mathbf{c}|\mathbf{x}_i|$, and thus the relationship between a dependent variable and the model is expressed as follows:

$$\mathbf{a}\mathbf{x}_i - \mathbf{c}\left|\mathbf{x}_i\right| \leq y_i + e_i\mathbf{c}\left|\mathbf{x}_i\right| \leq \mathbf{a}\mathbf{x}_i + \mathbf{c}\left|\mathbf{x}_i\right|. \tag{5}$$

As a result, LP of an interval type regression is re-expressed as follows:

$$\left.\begin{array}{ll} \min\limits_{\mathbf{a},\mathbf{c}} & \sum\limits_{i}^{n} W_i \\ \text{s.t.} & \mathbf{a}\mathbf{x}_i - \mathbf{c}\left|\mathbf{x}_i\right| \leq y_i + e_i\mathbf{c}\left|\mathbf{x}_i\right| \leq \mathbf{a}\mathbf{x}_i + \mathbf{c}\left|\mathbf{x}_i\right| \ \ (i = 1, 2, \ldots, n) \end{array}\right\} \tag{6}$$

Hereafter, the models obtained by Equations (3) and (6) are referred to as the conventional and improved models, respectively, to distinguish between the two interval regression models.

The model shown in Figure 2 is obtained when the improved model is applied to a numerical example of a bivariate distribution. The obtained conventional and improved models corresponding to $\mathbf{Y}_{e1}$ and $\mathbf{Y}_{e2}$, respectively, are as follows:

$$\mathbf{Y}_{e1} = (8.627, 6.717) + (0.634, 0.016)X, \tag{7}$$

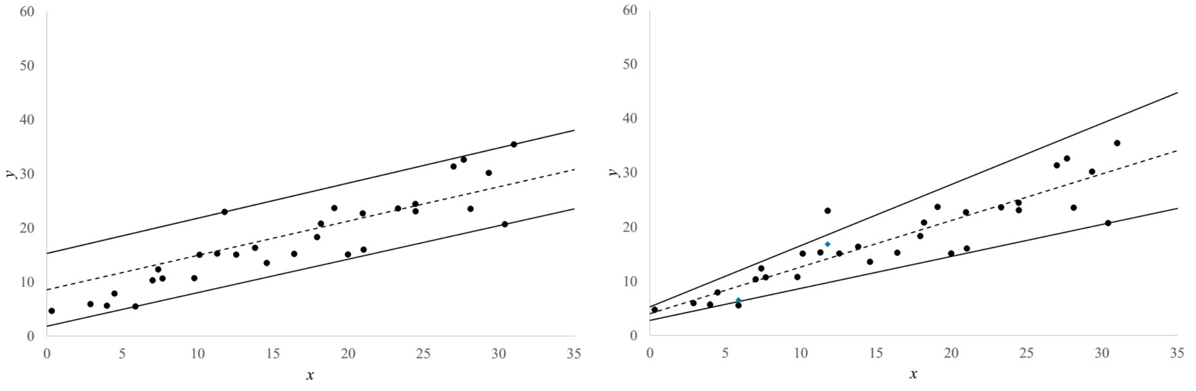
$$\mathbf{Y}_{e2} = (4.039, 1.244) + (0.859, 0.270)X. \tag{8}$$

Furthermore, the obtained least squares, Y_{es}, is as follows:

$$Y_{es} = 3.844 + 0.841X. \tag{9}$$

The sum of the widths of forecasted values is minimized by the objective function of LP, and thus the width of the model does not increase significantly from the origin. Hence, the centers of the model and the data distribution do not coincide in the conventional model, as shown in Figure 2(a). The improved model includes a slightly larger constant term and slope when compared to those of the least squares model. In this case, the centers of the model and data distribution appear to coincide with a slight decrease in the constant term and a slight increase in the slope.

A sample above the improved model shown in Figure 2(b) exists. A rhombus below the sample exists inside the possibility interval shown in the model. This corresponds to a dependent variable with the vagueness removed. Thus, the improved model processes samples that distort the shape of the model and illustrates the possibilities of the target system.



(a) Conventional model

(b) Improved model

FIGURE 2. Conventional and improved models in the numerical example

3. **Analysis of Leisure Time by Prefecture.** The Japanese government summarized various investigations with respect to individuals. The Ministry of Statistics published "*Statistical Observation of Prefecture 2017*" that systematically organized regional statistical data [12]. This section analyzes how prefectural citizens spend their leisure time by using data published in the report.

The data includes 111 variables in 18 sectors including population, households, natural environment, education, and labor. Prior to commencing the analysis, correlation coefficients were used and the 20 variables shown in Table 1 were selected. Volunteer activity, sports, and travel and excursion correspond to ways to spend leisure time. In this paper, the fore-mentioned three variables correspond to dependent variables, and the 17 attributes without high correlations are considered as independent variables to avoid multicollinearity. Additionally, standardized values are used to eliminate the influence of each variable unit.

Table 1. Prefecture data

Variable	Description
y_1	Annual participation rate of volunteer activities (ages 10 years and older)
y_2	Annual participation rate of sports (ages 10 years and older)
y_3	Annual participation rate of travel and excursions (ages 10 years and older)
x_1	Population
x_2	Dependency index (Proportion of the population aged 0-14 years and over 65/15-64 years)
x_3	Prefectural income per capita
x_4	Residential tax per capita
x_5	Number of elementary schools
x_6	Number of lower secondary schools
x_7	Number of upper secondary schools
x_8	Employed rate in primary industries with respect to all employed individuals
x_9	Employed rate in secondary industries with respect to all employed individuals
x_{10}	Employed rate in tertiary industries with respect to all employed individuals
x_{11}	Employment rate
x_{12}	Rate of outpatients
x_{13}	Number of new inpatients in a general hospital
x_{14}	Number of deaths due to lifestyle diseases
x_{15}	Medical expenditure per capita
x_{16}	Penal code criminal cases known to the police
x_{17}	Criminal cases involving theft that are known to the police

Variables were selected from 17 independent variables by using a stepwise method with Akaike's information criterion (AIC) while determining regression models. Conventional and improved models are determined by using the variables selected by the stepwise method. With respect to an interval model, a symmetrical triangular fuzzy number is used as regression coefficients.

The regression coefficients obtained by using the three dependent variables are shown in Table 2. Coefficients are compared in these tables, and coefficients that differ by 0.2 or more are shown in bold. Table 3 shows the features of the three models. The center of the interval models was used with respect to the residual sum of squares and the correlation coefficient. Additionally, the sum of the possibility grade with respect to the width of the

Table 2. The coefficients of the three models

		Least squares model	Conventional model		Improved model	
			a	**c**	**a**	**c**
y_1 (volunteer)	Const.	0.000	0.011	0.004	0.043	0
	x_2	0.921	0.753	0.350	0.833	0.251
	x_6	-0.541	-0.773	0.286	-0.641	0.232
	x_7	0.947	0.822	0.385	0.886	0.265
	x_{11}	0.507	**0.046**	0.074	0.327	0.161
	x_{12}	0.269	0.261	0.195	0.162	0.065
	x_{13}	-0.337	-0.195	0.327	-0.296	0.237
	x_{14}	-1.301	-1.444	0.000	-1.372	0.187
	x_{16}	-1.658	-1.678	0.350	**−1.390**	0.289
	x_{17}	1.559	1.378	0.298	1.487	0.285
y_2 (sports)	Const.	0.000	0.102	0.224	0.067	0.243
	x_2	0.763	0.771	0.029	0.811	0.125
	x_3	0.327	0.326	0.385	0.363	0.354
	x_{11}	-0.353	-0.433	0.109	-0.367	0.105
	x_{14}	-0.855	-0.925	0.564	-0.910	0.323
	x_{15}	-0.268	**−0.408**	0.020	-0.322	0.002
y_3 (travel)	Const.	0.000	0.012	0.000	-0.081	0.000
	x_4	0.391	0.358	0.405	0.415	0.317
	x_9	0.920	**1.120**	0.195	0.863	0.158
	x_{10}	0.523	**0.745**	0.039	0.350	0.261
	x_{12}	0.355	**0.531**	0.372	0.327	0.224
	x_{14}	-0.273	-0.279	0.051	-0.286	0.167
	x_{16}	-1.207	-1.216	0.111	-1.121	0.168
	x_{17}	1.169	1.256	0.160	**1.510**	0.197

Table 3. Features of the three models

	y_1 (volunteer)			y_2 (sports)			y_3 (travel)		
	LS	CV	IM	LS	CV	IM	LS	CV	IM
Residual sum of squares	18.55	48.05	36.21	11.93	14.91	12.49	11.08	15.65	15.77
Correlation coefficients	0.77	0.34	0.48	0.86	0.85	0.86	0.87	0.85	0.84
Sum of the possibility grades derived from the model and samples	–	23.40	24.77	–	25.50	25.02	–	23.32	23.78
Sum of the possibility grades to the widths of forecasted values	–	13.95	17.84	–	26.76	28.77	–	25.01	23.93
Sum of the widths of the forecasted values	–	170.42	146.44	–	100.69	87.67	–	95.90	96.94

LS: Least squares, CV: Conventional, IM: Improved.

forecasted values shown in Table 3 is obtained by $\sum_i \mu_i/W_i$, and thus an increase in the value improves the model.

The model involving volunteer activities uses the variable y_1 as a dependent variable. The residual sum of squares corresponds to the order of least squares, improved, and conventional models. As widely known, the least squares model defines coefficients to minimize the residual sum of squares. An interval model defines a coefficient to illustrate the possibility of the analysis object that possesses the minimum widths, and thus the center of the model does not show the feature of the analysis target. However, the least squares and the interval models can be compared. The residual sum of squares of the improved model approximately corresponds to 2.0 times that of the least squares model, and the residual sum of squares of the conventional model approximately corresponds to 2.6 times that of the least squares model. The sum of possibility grades and the sum of the grade to the forecasted value width of the improved model are approximately 1.7 and 3.9 higher than that of the conventional model respectively. Hence, the accuracy of the improved model is better with respect to the center of the model when compared with that of the conventional model. Furthermore, the width of the forecasted value of the improved model is also approximately 24.0 lower than that of the conventional model, and thus, the improved model is considered a good model. However, the correlation coefficient between the forecasted and the observed values is considerably small. It corresponds to approximately 0.44 times and 0.62 times that of the least squares model, respectively, for the conventional model and the improved model.

In the model related to sports, y_2 is used as a dependent variable. In the improved model, both the residual sum or squares and the correlation coefficients almost correspond to the same values as those of the least squares model. With respect to the width of the forecasted value, the improved model is approximately 13.0 smaller than the conventional one. Therefore, with respect to the sum of the possibility grade to the width of the forecasted value, the improved model is approximately 2.0 that of the conventional model. This indicates that the improved model illustrates the data distribution more accurately than the conventional model.

The model related to travel and excursions employs y_3 as a dependent variable. The residual sum of squares and the correlation coefficients reveals that the accuracy of the least squares model exceeds that of the interval model. The other indices are almost identical with respect to the interval model.

The characteristics of least squares and interval models are different as described above. Hence, a significant meaning to compare with respect to the statistical amount does not exist. However, it is considered that the interval models include the least squares and illustrate the possibility of the analysis target while comparing each coefficient. A feature common to the three dependent variables is that the constant terms are not considered and the signs of each variable are the same. Therefore, based on the magnitude of each statistical value, it is possible to determine the number of participants for volunteer activities, sports, and travel and excursions by using the prefecture.

With respect to volunteer activities, there are many volunteer participants despite the existence of several dependent individuals and individuals in an upper secondary school or a high employment rate. Conversely, the participation in volunteer activities is low given the existence of several lower secondary school individuals and inpatients. Additionally, it is interesting to note that that volunteer participation tends to decrease when several deaths are caused due to lifestyle diseases and penal code criminal cases.

With respect to sports activities, the number of participants decreases if there are many deaths due to lifestyle diseases. Appropriate exercises correspond to the prevention of lifestyle diseases, and thus this is understood as a matter of course. Additionally, the proportion of sports activities increases if the dependent population is high. Furthermore, the participation rate in sports activities is decreased by an increase in the employment rate and increased by a prefectural income per capita. The population of sports is also high when the national medical expenses per capita are small.

With respect to travel and excursions, it is interesting that penal code criminal cases decrease the participation rate, and the number of criminal cases involving theft increases the same. Increase in the employment rate of secondary and tertiary industries and residential tax increases the participant number for travel and excursions. Additionally, fewer individuals engage in travel and excursions if the deaths due to lifestyle diseases are high.

The reduction of activities involving national medical expenses, deaths involving lifestyle diseases, employment rate, and number of penal code criminal cases correspond to effective methods to spend leisure time in the fore-mentioned three cases. Additionally, prefectural income and residential tax indicating economic situation show economic margin, and this can increase both activities.

4. **Conclusions.** In this paper, ways citizens spend their leisure time by using statistical data showing characteristics of prefectures are analyzed. A regression analysis using prefecture-specific data is performed to provide a relative comparison that emphasizes the difference between prefectures. Based on this, it is confirmed that the following indicators influence the number of participants in leisure activities.

- A decrease in leisure activities results from increased national medical expenses, patients with lifestyle diseases, employment rate, and the number of penal code criminal cases.
- An increase in leisure activities results from increased prefectural income and resident tax.

In the data set on the number of volunteer participants, optimal results are not obtained by checking the statistical accuracy by using the center of the interval models. However, the interval models illustrate the possibility of analysis targets, and thus issues do not exist with respect to both the conventional and improved models. Additionally, the results confirmed that the obtained models are appropriate. In the data set on sports activities, the improved model exhibited an extremely high forecasting accuracy. An interval model is determined by the LP problem, and thus the forecast accuracy significantly changes based on the objective function and constraints. In this paper, effective models are obtained without adjusting the LP problem, and this confirmed that the improved model is a useful model.

A future study should employ the model for verification purposes by using the data real system with an individual as a constitute element and improve it to suggest a more practical model.

REFERENCES

[1] R. Coppi, P. D'Urso, P. Giordani and A. Santoro, Least squares estimation of a linear regression model with LR fuzzy response, *Computational Statistics & Data Analysis*, vol.51, no.1, pp.267-286, 2006.

[2] P. Diamond, Fuzzy least squares, *Information Science*, vol.46, no.3, pp.141-157, 1988.

[3] P. Diamond, Least squares and maximum likelihood regression for fuzzy linear models, in *Fuzzy Regression Analysis*, J. Kacprzyk and M. Fedrizzi (eds.), Omnitech Press, 1992.

[4] P. D'Urso and T. Gastaldi, A least-squares approach to fuzzy linear regression analysis, *Computational Statistics & Data Analysis*, vol.34, no.4, pp.427-440, 2000.

[5] P. Guo and H. Tanaka, Fuzzy DEA: A perceptual evaluation method, *Fuzzy Sets and Systems*, vol.119, no.1, pp.149-160, 2001.

[6] D. H. Hong, C. Hwang and C. Ahn, Ridge estimation for regression models with crisp inputs and Gaussian fuzzy output, *Fuzzy Sets and Systems*, vol.142, no.2, pp.307-319, 2004.

[7] M. Inuiguchi and T. Tanino, Interval linear regression methods based on Minkowski difference: A bridge between traditional and interval linear regression models, *Kybernetika*, vol.42, no.4, pp.423-440, 2006.

[8] H. Lee and H. Tanaka, Upper and lower approximation models in interval regression using regression quantile techniques, *European Journal of Operational Research*, vol.116, no.3, pp.653-666, 1999.

[9] M. Modarres, E. Nasrabadi and M. M. Nasrabadi, Fuzzy linear regression model with least square errors, *Applied Mathematics and Computation*, vol.163, no.2, pp.977-989, 2005.

[10] A. A. Ramli, J. Watada and W. Pedrycz, Real-time fuzzy regression analysis: A convex hull approach, *European Journal of Operational Research*, vol.210, no.3, pp.606-617, 2011.

[11] A. A. Ramli, J. Watada and W. Pedrycz, A combination of genetic algorithm-based fuzzy c-means with a convex hull-based regression for real-time fuzzy switching regression analysis: Application to industrial intelligent data analysis, *IEEJ Trans. Electrical and Electronic Engineering*, vol.9, no.1, pp.71-82, 2014.

[12] Statistics Bureau, Ministry of Internal Affairs and Communications, *Statistical Observation of Prefecture 2017*, 2017.

[13] H. Tanaka and J. Watada, Possibilistic linear systems and their application to the linear regression model, *Fuzzy Sets and Systems*, vol.27, no.3, pp.275-289, 1988.

[14] J. Watada and W. Pedrycz, A fuzzy regression approach to acquisition of linguistic rules, in *Handbook of Granular Computing*, W. Pedrycz, A. Skowron and V. Kreinovich (eds.), Wiley, 2008.

[15] Y. Yabuuchi and J. Watada, Model building based on central position for a fuzzy regression model, *Proc. of Czech-Japan Seminar 2006*, pp.114-119, 2006.

[16] Y. Yabuuchi and J. Watada, Fuzzy regression model building through possibility maximization and its application, *ICIC Express Letters*, vol.4, no.2, pp.505-510, 2010.

[17] Y. Yabuuchi and J. Watada, Fuzzy robust regression model by possibility maximization, *JACIII*, vol.15, no.4, pp.479-484, 2011.

[18] Y. Yabuuchi, Japanese economic analysis by a fuzzy regression model building through possibility maximization, *Proc. of the 6th International Conference on Soft Computing and Intelligent Systems, and the 13th International Symposium on Advanced Intelligent Systems*, pp.1772-1777, 2012.

[19] Y. Yabuuchi and J. Watada, Fuzzy robust regression model building through possibility maximization and analysis of Japanese major rivers, *ICIC Express Letters*, vol.9, no.4, pp.1033-1041, 2015.

[20] Y. Yabuuchi, Centroid-based fuzzy regression model, *ICIC Express Letters*, vol.9, no.12, pp.3299-3306, 2015.

[21] Y. Yabuuchi, Possibility grades with vagueness in fuzzy regression models, *Procedia Computer Science*, vol.112, pp.1470-1478, 2017.