Fuzzy data analysis based on regression modeling

N V Sapkina
Department of Applied Mathematics, Informatics and Mechanics, Voronezh State University, 1, University Square, Voronezh, 394018, Russia
E-mail: natashasapkina@yandex.ru

Abstract. This article considers the data-mining system based on fuzzy regression modeling, which allows to extract useful previously unknown patterns such as groups of data records, unusual records and dependencies. Using the fuzzy least-squares approach, we describe the estimation of fuzzy linear regression model with $LR$-type fuzzy parameters. In conclusion, the received theoretical results are used to analyze experimental data.

1. Introduction
Use of computer-based technologies reveals the importance of tasks related to the information processing for the data extraction. It becomes more and more popular to create data-mining systems, which allow to extract useful previously unknown patterns such as groups of data records, unusual records and dependencies based on classification, clustering, statistical analysis, association rule mining and other methods. If experimental data is presented in the form of statistical series or can be generated from a database, then we can use ordinary regression modeling techniques for data analysis, assuming that, the data is numerical. However, if we consider fuzzy data, then traditional regression analysis methods are not applicable and should be modified under the fuzzy environment.

In order to consider the fuzziness in regression analysis, Tanaka et al. (1982) first proposed a study of fuzzy linear regression model. They considered the parameter estimations of such models under two factors, namely the degree of the fitting and the vagueness of the model. The estimation problems were then transformed into linear programming based on these two factors. Since the measure of best fit by residuals under fuzzy consideration is not presented in Tanakas approach, Diamond (1988) proposed the fuzzy least-squares approach, which is a fuzzy extension of the ordinary least squares based on a new defined distance on the space of fuzzy numbers. In general, the fuzzy regression methods can be roughly divided into two categories. The first is based on Tanakas linear programming approach. The second category is based on the fuzzy least-squares approach.

In section 2, we introduced the fuzzy number and its operation, a simple distance formula, fuzzy linear regression model and its least squares estimate, data-mining system.

In section 3, we used the theoretical results in the previous chapter to analyze experimental data.
2. Materials and methods

2.1. LR-type fuzzy numbers and their operations

Let $\mathcal{R}$ be a one-dimensional Euclidean space with its norm denoted by $\| \cdot \|$. A fuzzy number is an upper semicontinuous convex function $F : \mathcal{R} \to [0, 1]$, where $\{ x \in \mathcal{R} | F(x) = 1 \}$ is non-empty. In other words, a fuzzy number $A$ is defined as a convex normalized fuzzy set of the real line $\mathcal{R}$ so that there exists exactly one $x_0 \in \mathcal{R}$ with $F(x_0) = 1$ and its membership function $F(x)$ is piecewise continuous [1].

Definition 1.

Let $L$ (and $R$) be decreasing, shape functions from $\mathcal{R}^+ \to [0, 1]$ with $L(0) = 1$; $L(x) < 1$ for all $x > 0$; $L(x) > 0$ for all $x < 1$; $L(1) = 0$ for all $x$ and $L(+\infty) = 0$. Then a fuzzy number $A$ is called as LR-type fuzzy number, if for $\mu, \alpha > 0$, $\beta > 0$ in $\mathcal{R}$ its membership function is defined as [2]:

$$m_a(x) = \begin{cases} L \left( \frac{\mu - x}{\alpha} \right), & \text{if } x \leq \mu, \\ R \left( \frac{x - \mu}{\beta} \right), & \text{if } x \geq \mu. \end{cases}$$

where $\mu$ — is called the center (mean or mode) value of $A$, $\alpha$ and $\beta$ — are called left and right spreads, respectively. Symbolically, $A$ is denoted by $(\mu, \alpha, \beta)$. If $\alpha = \beta$, then $A$ is called symmetrical LR-type fuzzy number, denoted by $A = (\mu, \alpha)$ [3]. For instance, the algebraic and geometric characteristics of the membership function of the more utilized LR-type fuzzy number, the triangular fuzzy number, are shown in the following:

$$m_a(x) = \begin{cases} 1 - \frac{\mu - x}{\alpha}, & \text{if } \mu - \alpha \leq x \leq \mu, \\ 1 - \frac{x - \mu}{\beta}, & \text{if } \mu \leq x \leq \mu + \beta. \end{cases}$$

Definition 2. Let $A = (\mu_a, \alpha_a, \beta_a)$ and $B = (\mu_b, \alpha_b, \beta_b)$ be two LR-type fuzzy numbers. Then by the extension principle, the following operations are defined [4]:

1. $A + B = (\mu_a + \mu_b, \alpha_a + \alpha_b, \beta_a + \beta_b)$.
2. $\lambda A = (\lambda \mu_a, \lambda \alpha_a, \lambda \beta_a)$, $\lambda > 0$ and $\lambda A = (\lambda \mu_a, -\lambda \beta_a, -\lambda \alpha_a)$, $\lambda < 0$.
3. $A - B = (\mu_a - \mu_b, \alpha_a + \beta_b, \alpha_b + \beta_a)$.

Definition 3 (a Euclidean distance formula).

Let $A = (\mu_a, \alpha_a, \beta_a)$ and $B = (\mu_b, \alpha_b, \beta_b)$ be two LR-type fuzzy numbers, then the distance between them is defined as [5]:

$$D = \sqrt{(\mu_a - \mu_b)^2 + (\alpha_a - \alpha_b)^2 + (\beta_a - \beta_b)^2}.$$ 

Definition 4 (a $\gamma$-cut formula).

Let $A = (\mu_a, \alpha_a, \beta_a)$ be LR-type fuzzy number and $\gamma \in (0, 1)$, then the $\gamma$-cut or $\gamma$-level is the crisp set

$$\sigma_\gamma(A) = \{ x : m_a(x) \geq \gamma, x \in X \}.$$ 

The strong $\gamma$-cut is the crisp set

$$\omega_\gamma(A) = \{ x : m_a(x) > \gamma, x \in X \}.$$ 

For example, to find the $\gamma$-cut of triangular fuzzy number $A$, we first set $\alpha \in (0, 1]$ to both left and right reference functions of $A$. That is, $\gamma = 1 - (\mu_a - x)/\alpha_a$ and $\gamma = 1 - (x - \mu_a)/\beta_a$. Expressing $x$ in terms of $\gamma$ we have $x = \mu_a - \alpha_a(1 - \gamma)$ and $x = \mu_a + \beta_a(1 - \gamma)$. The $\gamma$-cut of $A$ is defined as:

$$\sigma_\gamma(A) = [\mu_a - \alpha_a(1 - \gamma), \mu_a + \beta_a(1 - \gamma)].$$
2.2. General fuzzy linear regression model

Suppose that observations consist of data pairs \( \{(x_i, Y_i)\}_{i=1}^{\infty} \), where \( x_i \in \mathcal{R} \) and \( Y_i = (\mu_{yi}, \alpha_{yi}, \beta_{yi}) \) are LR-type fuzzy numbers. Consider the following general fuzzy linear regression model [6]:

\[
Y_i = A_0 + A_1x_{i1} + A_2x_{i2} + \cdots + A_nx_{in} + E_i, \quad i = 1, m,
\]

where \( x_{ik} \in \mathcal{R}, \ k = 1, n, \ Y_i = (\mu_{yi}, \alpha_{yi}, \beta_{yi}), \ i = 1, m, \) are LR-type fuzzy numbers, \( \mu_{yi} \) is the center (mean or mode) value, \( \alpha_{yi} \) and \( \beta_{yi} \) are left and right spreads, respectively; \( A_j = (\mu_{aj}, \alpha_{aj}, \beta_{aj}), \ j = 0, n, \) are the fuzzy regression parameters, which have the same membership function as \( Y_i; \ E_i = (\mu_{ei}, \alpha_{ei}, \beta_{ei}), \ i = 1, m, \) are random errors, LR-type fuzzy numbers.

Based on the definition 3, we can use ordinary least-squares method to estimate the fuzzy parameters \( A_j, \ j = 0, n, \) in the general fuzzy linear regression model (1) [7].

Assuming that \( Y_i = (\mu_{yi}, \alpha_{yi}, \beta_{yi}), \ i = 1, m, \) and \( A_j = (\mu_{aj}, \alpha_{aj}, \beta_{aj}), \ j = 0, n, \) have the same membership function, after appropriate translation, we can make all of \( x_{ik} > 0, \ i = 1, m, k = 1, n. \)

Estimation model for (1) can be expressed as

\[
\tilde{Y}_i = \tilde{A}_0 + \tilde{A}_1x_{i1} + \cdots + \tilde{A}_nx_{in}, \quad i = 1, m,
\]

where \( \tilde{Y}_i = (\tilde{\mu}_{yi}, \tilde{\alpha}_{yi}, \tilde{\beta}_{yi}), \ i = 1, m \) — estimations of \( Y_i, \ \tilde{A}_j = (\tilde{\mu}_{aj}, \tilde{\alpha}_{aj}, \tilde{\beta}_{aj}), \ j = 0, n \) — estimations of model parameters [6].

Accepting that all \( x_{ik} > 0, \ i = 1, m, k = 1, n, \) we can rewrite (2) as follows:

\[
(\tilde{\mu}_{yi}, \tilde{\alpha}_{yi}, \tilde{\beta}_{yi}) = (\tilde{\mu}_{a0}, \tilde{\alpha}_{a0}, \tilde{\beta}_{a0}) + (\tilde{\mu}_{a1}, \tilde{\alpha}_{a1}, \tilde{\beta}_{a1}) x_{i1} + \cdots + (\tilde{\mu}_{an}, \tilde{\alpha}_{an}, \tilde{\beta}_{an}) x_{in}, \quad i = 1, m.
\]

According to the Euclidean distance formula of definition 3, the least-squares estimates of \( \mu_{aj}, \alpha_{aj} \) and \( \beta_{aj} \), are the values of \( \tilde{\mu}_{aj}, \tilde{\alpha}_{aj} \) and \( \tilde{\beta}_{aj} \), which minimize the value of \( D^2 \), where

\[
D^2 = \sum_{i=1}^{m} \left[ \frac{(\mu_{yi} - (\tilde{\mu}_{a0} + \tilde{\mu}_{a1}x_{i1} + \cdots + \tilde{\mu}_{an}x_{in}))^2}{2} + \frac{(\alpha_{yi} - (\tilde{\alpha}_{a0} + \tilde{\alpha}_{a1}x_{i1} + \cdots + \tilde{\alpha}_{an}x_{in}))^2}{2} + \frac{(\beta_{yi} - (\tilde{\beta}_{a0} + \tilde{\beta}_{a1}x_{i1} + \cdots + \tilde{\beta}_{an}x_{in}))^2}{2} \right].
\]

Let \( ||\vec{v}|| \) denote the length of vector \( \vec{v} \), then by using vector and matrix expressions \( D^2 \) can be rewritten as

\[
D^2 = ||X\tilde{\mu}_a - \mu_y||^2 + ||X\tilde{\alpha}_a - \alpha_y||^2 + ||X\tilde{\beta}_a - \beta_y||^2,
\]

where \( X \) is a \( m \times (n + 1) \) design matrix, \( \tilde{\mu}_a = (\tilde{\mu}_{a0}, \tilde{\mu}_{a1}, \cdots, \tilde{\mu}_{an})^T, \tilde{\alpha}_a = (\tilde{\alpha}_{a0}, \tilde{\alpha}_{a1}, \cdots, \tilde{\alpha}_{an})^T, \tilde{\beta}_a = (\tilde{\beta}_{a0}, \tilde{\beta}_{a1}, \cdots, \tilde{\beta}_{an})^T, \)

\( \tilde{\mu}_y = (\tilde{\mu}_{y1}, \cdots, \tilde{\mu}_{ym}) \), \( \tilde{\alpha}_y = (\tilde{\alpha}_{y1}, \cdots, \tilde{\alpha}_{ym}) \) and \( \tilde{\beta}_y = (\tilde{\beta}_{y1}, \cdots, \tilde{\beta}_{ym}) \). Then we obtain the following equations [6]:

\[
\begin{align*}
\frac{\partial D^2}{\partial \tilde{\mu}_a} &= 2X^TX\tilde{\mu}_a - 2X^T\mu_y = 0, \\
\frac{\partial D^2}{\partial \tilde{\alpha}_a} &= X^TX\tilde{\alpha}_a - X^T\alpha_y = 0, \\
\frac{\partial D^2}{\partial \tilde{\beta}_a} &= X^TX\tilde{\beta}_a - X^T\beta_y = 0.
\end{align*}
\]

The solutions of \( \tilde{\mu}_a, \tilde{\alpha}_a \) and \( \tilde{\beta}_a \), which minimize \( D^2 \), can be expressed as

\[
\begin{align*}
\tilde{\mu}_a &= (X^TX)^{-1}X^T\mu_y, \\
\tilde{\alpha}_a &= (X^TX)^{-1}X^T\alpha_y, \\
\tilde{\beta}_a &= (X^TX)^{-1}X^T\beta_y.
\end{align*}
\]
The above method used regression with respect to center and spreads. The estimation results are not related to the membership functions. But, in real data analyses, this method provides good results in the estimation of fuzzy parameter values.

In order to estimate the quality of the constructed fuzzy linear regression model, we can apply traditional methods of assessing model fit as well as the accuracy evaluation [8].

A correlation coefficient is useful in assessing model fit and determines the strength of relationship between the dependent and independent variables. Based on the traditional linear regression analysis, we can define a new formula of correlation coefficient under the fuzzy environment:

$$ R = (\mu_r, \alpha_r, \beta_r), $$

where $\mu_r$, $\alpha_r$, $\beta_r$ are as follows:

$$ \mu_r = \frac{n \sum_{i=1}^{n} x_i \mu_{yi} - \left( \sum_{i=1}^{n} x_i \right) \left( \sum_{i=1}^{n} \mu_{yi} \right)}{\sqrt{n \sum_{i=1}^{n} x_i^2 - \left( \sum_{i=1}^{n} x_i \right)^2} \cdot \sqrt{n \sum_{i=1}^{n} \mu_{yi}^2 - \left( \sum_{i=1}^{n} \mu_{yi} \right)^2}}. $$

$$ \alpha_r = \frac{n \sum_{i=1}^{n} x_i \alpha_{yi} - \left( \sum_{i=1}^{n} x_i \right) \left( \sum_{i=1}^{n} \alpha_{yi} \right)}{\sqrt{n \sum_{i=1}^{n} x_i^2 - \left( \sum_{i=1}^{n} x_i \right)^2} \cdot \sqrt{n \sum_{i=1}^{n} \alpha_{yi}^2 - \left( \sum_{i=1}^{n} \alpha_{yi} \right)^2}}. $$

$$ \beta_r = \frac{n \sum_{i=1}^{n} x_i \beta_{yi} - \left( \sum_{i=1}^{n} x_i \right) \left( \sum_{i=1}^{n} \beta_{yi} \right)}{\sqrt{n \sum_{i=1}^{n} x_i^2 - \left( \sum_{i=1}^{n} x_i \right)^2} \cdot \sqrt{n \sum_{i=1}^{n} \beta_{yi}^2 - \left( \sum_{i=1}^{n} \beta_{yi} \right)^2}}. $$

Using the definition 4, the $\gamma$-cut of LR-type fuzzy number $R = (\mu_r, \alpha_r, \beta_r)$ can be expressed as

$$ [R(\gamma), R(\gamma)] = [\mu_r - \alpha_r(1 - \gamma), \mu_r + \beta_r(1 - \gamma)], \quad \gamma \in [0, 1]. $$

The crisp correlation coefficient is defined in this case as follows:

$$ \varepsilon \in [R(\gamma), R(\gamma)] \Rightarrow r = R(\gamma) \cdot \varepsilon + R(\gamma) \cdot (1 - \varepsilon), \quad \gamma \in [0, 1]. $$

The closer the coefficient is to either $-1$ or $1$, the stronger the correlation between the variables $Y$ and $x$. A large value of $r$ is also indicates the good quality of the constructed fuzzy linear regression model.

Based on (1) and (2), we can use the ordinary accuracy evaluation to estimate the average relative error under the fuzzy environment:

$$ E_{rel} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - \bar{Y}_i}{Y_i} \right| \cdot 100\%. $$

Using the definition 4, the $\gamma$-cut of LR-type fuzzy numbers $Y_i = (\mu_{yi}, \alpha_{yi}, \beta_{yi})$ and $\bar{Y}_i = (\tilde{\mu}_{yi}, \tilde{\alpha}_{yi}, \tilde{\beta}_{yi})$ can be expressed as:

$$ [\sum_i(\gamma), \underline{Y}_i(\gamma)] = [\mu_y - \alpha_y(1 - \gamma), \mu_y + \beta_y(1 - \gamma)], \quad \gamma \in [0, 1]. $$
\[
\left[ \widetilde{Y}_i(\gamma), \bar{Y}_i(\gamma) \right] = [\tilde{\mu}_y - \tilde{\alpha}_y(1 - \gamma), \tilde{\mu}_y + \tilde{\beta}_y(1 - \gamma)], \quad \gamma \in [0, 1].
\]

The crisp average relative error is defined as follows:
\[
\varepsilon \in \left[ \Sigma_i(\gamma), \bar{Y}_i(\gamma) \right] \Rightarrow \lambda_i^\gamma = \Sigma_i(\gamma) \cdot \varepsilon + \bar{Y}_i(\gamma) \cdot (1 - \varepsilon), \quad \gamma \in [0, 1],
\]
\[
\varepsilon \in \left[ \Sigma_i(\gamma), \bar{Y}_i(\gamma) \right] \Rightarrow \lambda_i^\gamma = \Sigma_i(\gamma) \cdot \varepsilon + \bar{Y}_i(\gamma) \cdot (1 - \varepsilon), \quad \gamma \in [0, 1].
\]

As a result, we obtain the following formula of crisp average relative error:
\[
e_{rel} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\lambda_i^\gamma - \lambda_i^\gamma}{\lambda_i^\gamma} \right| \cdot 100 \%, \quad \gamma \in [0, 1].
\]

The closer the average relative error to 0, the fitter the constructed fuzzy linear regression model. If \( e_{rel} > 15\% \), then the above method of estimating the regression parameters provides bad quality. A value of \( e_{rel} \leq 5\% \) indicates good results in the estimation of fuzzy parameter values.

2.3. Data-mining system

The developed data-mining system is based on the information repository technology [9]. This system processes big amount of data, performs computer assisted retrieval and allows to extract useful previously unknown patterns such as groups of data records, unusual records and dependencies. The analysis results can further be applied for making management-related decisions and improving business processes.

The structure of data-mining system (see figure 1) has two applications — analytical application (main) and administration system (auxiliary application). The auxiliary application is used for execution of SQL-queries with the participation of analyst.

Information about forms and transitions is contained in a special database, which can be local or remote. Figure 2 shows information repository structure based on entities tabs, specific pieces of information.

Entity "Indicator" contains descriptions of economical, technical and other indicators that are necessary for conducting analytical work. They have a hierarchical structure. Entity "Unit of measure" is used to store the data in unified form and contains information about the measurements, the conversion factor between them is represented by attribute Coef (Multiplier). Entity "Data" contains the values of the indicators for each dimension in the form of three attributes: for mean value, left and right spreads of \( LR \)-type fuzzy numbers. In addition, it refers to the identifier of the indicator, its type, the validity period and the actual date, which allows, if necessary, to have information about the values of certain indicators in a more detailed form.

Administration system includes the following tasks:
- Authorization of users;
- User management;
- Metadata management;
- Validity periods management;
- Uploading data to information repository.

The user authorization function stands for authentification of users, who are registered in the system and have access to work.

The user management function is designed to create, modify and delete a user, view the list of all users of the system, whose basic data should correspond to the "User" entity.

The metadata management function is used to manage entities "Unit of measure" and "Indicator" and allows to work with a hierarchical data structure, dividing units of measure
and indicators into base and derived ones. It also provides the ability to add, edit and delete data entries.

The period control function is designed to create new reporting periods and close old ones. The data must correspond to the entity "Validity period".

The uploading data function stands for entering data into information repository by available list of indicators and the current validity period.

The input data for the described administration system is the data from entities "User", "Unit of measure", "Indicator" and "Validity period". The output data is the number of uploaded records to information repository, as well as the number of records for a particular reporting period.

3. Results and discussion
In this section, we use experimental data to illustrate the theoretical results which we obtained in the previous sections. The developed data-mining system is applied to analyze the quality of produced paints and varnishes and for making management-related decisions in order to improve business processes. The mining process consists of the following steps:
1. Data preparation.
   The data set contains three independent variables \( x_1 \) — filler, %, \( x_2 \) — lacquer solvent, %, and \( x_3 \) — binder, %, one fuzzy response variable (quality of produced paints and varnishes) and twelve data points (see table 1). We consider LR-type fuzzy response values \( Y_i = (\mu_{yi}, \alpha_{yi}, \beta_{yi}) \).

| \( x_1 \) |  11  |  10  |  11  |  12  |  13,5 |  14  |  15  |  16  |  17  |  17,5 |  19  |  20  |
|----------|------|------|------|------|-------|------|------|------|------|-------|------|------|
| \( x_2 \) |  10  |  10,5|  12,5|  12  |  13    |  13,5|  14  |  16  |  16,5|  17    |  18   |  20   |
| \( x_3 \) |  12  |  12,5|  13  |  14,5|  16    |  16,5|  17  |  18  |  19,5|  20,5  |  21   |  22   |
| \( \mu_y \) |  11,2 |  12,5|  12,9|  14,1|  14,8  |  16   |  17,5|  18,9|  18,9|  20    |  21,1 |  22,2 |
| \( \alpha_y \) |  0,9  |  0,1  |  1   |  0,5  |  1,1   |  0,1  |  0,5  |  0,3  |  1,3  |  1,1   |  0,2  |
| \( \beta_y \) |  0,2  |  0,4  |  0,5  |  1,1  |  0,7   |  0,1  |  0,08 |  1,2  |  0,7  |  0,48  |  1,9   |  0,4  |

By using the autoassociative neural network with one middle layer neuron, we obtain the contracted representation of independent variables in vector form [10]: \( x = (11, 11.5, 13, 14, 15, 15.5, 16, 17, 18, 19, 21, 22) \).

2. Fuzzy linear regression modeling.
   Based on the proposed least-squares method of estimating the fuzzy parameters, the fuzzy linear regression model is constructed as: \( \tilde{Y} = (0, 0.6, 0.6, 0.01) + (1, 0.004, 0.04)x \) (see table 2).

| \( \tilde{\mu}_y \) |  11,5 |  12,1 |  13,6 |  14,6 |  15,6 |  16,1 |  16,6 |  17,6 |  18,6 |  19,6 |  21,6 |  22,6 |
| \( \tilde{\alpha}_y \) |  0,65  |  0,65  |  0,65  |  0,66  |  0,66  |  0,66  |  0,67  |  0,67  |  0,68  |  0,68  |  0,68  |  0,69  |
| \( \tilde{\beta}_y \) |  0,45  |  0,47  |  0,53  |  0,57  |  0,61  |  0,63  |  0,66  |  0,69  |  0,73  |  0,77  |  0,85  |  0,89  |

By using the formula of (7), we can define the correlation coefficient under the fuzzy environment: \( R = (0.68, 0.02, 0.4) \). The \( \gamma \)-cut of LR-type fuzzy number \( R = (0.68, 0.02, 0.4) \) can be expressed as \( [R(\gamma), \overline{R}(\gamma)] = [0.68 - 0.02 \cdot (1 - \gamma), 0.68 + 0.4 \cdot (1 - \gamma)] \), \( \gamma \in [0, 1] \), then the crisp correlation coefficient is defined in this case as follows:

\[
    r = \frac{R(\gamma) \cdot \varepsilon + \overline{R}(\gamma) \cdot (1 - \varepsilon)}{2}, \quad \varepsilon = \frac{R(\gamma) + \overline{R}(\gamma)}{2}, \quad \gamma \in [0, 1].
\]

If \( \gamma \in [0, 1] \), then \( r \in [0.68, 0.87] \), which indicates the good quality of the constructed fuzzy linear regression model.

3. Accuracy evaluation.
   Based on experimental data and constructed FLR model \( \tilde{Y} = (0.6, 0.6, 0.01) + (1, 0.004, 0.04)x \), we can use the ordinary accuracy evaluation to estimate the average relative error under the fuzzy environment:

\[
    \varepsilon = (\mu_\varepsilon, \alpha_\varepsilon, \beta_\varepsilon) = \left( \frac{1}{12} \sum_{i=1}^{12} (\tilde{\mu}_{yi} - \mu_{yi}), \frac{1}{12} \sum_{i=1}^{12} (\tilde{\alpha}_{yi} - \alpha_{yi}), \frac{1}{12} \sum_{i=1}^{12} (\tilde{\beta}_{yi} - \beta_{yi}) \right) = (0.4, 0.4, 0.3). \]
Using the definition 4, the \( \gamma \)-cut of LR-type fuzzy numbers \( Y_i = (\mu_{yi}, \alpha_{yi}, \beta_{yi}) \) and \( \tilde{Y}_i = (\tilde{\mu}_{yi}, \tilde{\alpha}_{yi}, \tilde{\beta}_{yi}) \) can be expressed as: 

\[
\left[ \mu_{yi} - \alpha_{yi}(1 - \gamma), \mu_{yi} + \beta_{yi}(1 - \gamma) \right] \quad \text{and} \quad \left[ \tilde{\mu}_{yi} - \tilde{\alpha}_{yi}(1 - \gamma), \tilde{\mu}_{yi} + \tilde{\beta}_{yi}(1 - \gamma) \right],
\]

respectively (see figures 3, 4, 6).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{The \( \gamma \)-cut of crisp and LR-type fuzzy numbers.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{The \( \gamma \)-cut of crisp and LR-type fuzzy numbers.}
\end{figure}

Figure 7 shows the average relative error, where 

\[\bar{e} = \left( \frac{1}{12} \sum_{i=1}^{12} |Y - \tilde{Y}| \right) / 12,\]

depends on the \( \gamma \)-cut. We can define, that the more the value of the \( \gamma \)-cut, the lower the average relative error. A value of \( \bar{e} = 5.5 \% \) indicates good results in the estimation of fuzzy parameter values.
Figure 5. The 0,5-cut of LR-type fuzzy number.

Figure 6. The 0,5 γ-cut of crisp and LR-type fuzzy numbers.

Figure 7. The average relative error depends on the γ-cut.
4. Interpretation of analysis.

The developed data-mining system shows good results and can be applied to analyze the quality of produced paints and varnishes and for making management-related decisions and improving business processes.

4. Conclusion

In this article, we considered the data-mining system based on fuzzy regression modeling, which allows to extract useful previously unknown patterns such as groups of data records, unusual records and dependencies. Using the fuzzy least-squares approach, we proposed the estimation of fuzzy linear regression model with $LR$-type fuzzy parameters. The received theoretical results were used to analyze experimental data and shows, that the developed data-mining system indicates good results and can be applied to analyze the quality of produced paints and varnishes and for making management-related decisions and improving business processes.

References

[1] Hong D H 2004 Fuzzy Sets and Systems 142 307–319
[2] Zimmermann H J 1997 Fuzzy Set Theory and its Applications (Kluwer Academic Publishers)
[3] Pegat A 2009 Fuzzy modeling and control (Binom)
[4] Dubois D 1980 Fuzzy Sets and Systems: Theory and Applications (Academic Press)
[5] Tanaka H 1987 Fuzzy Sets and Systems 24 363–375
[6] Ledeneva T and Sapkina N 2011 Modern economics: Problems and solutions 10 174–181
[7] Yang M S 2002 Fuzzy Sets and Systems 126 389–399
[8] Kremer N 2002 Econometrics (Unity)
[9] Trofimov V 2013 Information systems and technologies in economics and management (Yurayt)
[10] Sapkina N 2013 Management systems and information technologies 4 27–30