Computers in Biology and Medicine ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Analysis of the relation between health statistics and eating habits in Japanese prefectures using fuzzy robust regression model

Yoshiyuki Yabuuchi [a,*], Takayuki Kawaura [b], Junzo Watada [c]

[a] Shimonoseki City University, 2-1-1 Daigaku-cho, Shimonoseki, Yamaguchi 751-8510, Japan
[b] Kansai Medical University, 2-5-1 Shin-machi, Hirakata, Osaka 573-1010, Japan
[c] Waseda University, 2-7 Hibikino, Wakamatsuku, Kitakyushu, Fukuoka 808-013, Japan

ARTICLE INFO

Article history:
Received 5 February 2015
Accepted 11 January 2016

Keywords:
Medical care cost
Number of outpatients
Lifestyle-related disease
Fuzzy robust regression model
Possibility grade

ABSTRACT

In recent years, the Japanese Ministry of Health, Labour, and Welfare is working to improve citizen's lifestyle and social environment to improve their health. This is because of the following reasons. Diseases related to lifestyle such as malignant neoplasms, heart disease, and cerebrovascular disease account for about 60% of the deaths in 2013. In addition, 32% of all medical expenditures are made on lifestyle-related disease. Lifestyle-related diseases can be prevented by daily exercise, a well-balanced diet, and not smoking. This ministry is promoting measures such as dietary education, physical activity, and exercise. Improvement of diet is the easiest way to reduce the occurrence of lifestyle-related diseases. Thus, in this paper, we analyze the relation between health and diet using our fuzzy robust regression model.

1. Introduction

As of July 1, 2014, the population of Japan was 127 million [15]. The population pyramid is distorted, with children constituting a smaller proportion of the population than older people do. The population has decreased by 0.21 million as compared to the previous year. The population under 15 years of age is 16 million, and it has decreased by 0.17 million compared with the previous year. Similarly, the population above 65 years of age is 33 million, and it has increased by 1.09 million. As aging in Japan proceeds, the structure of the population pyramid will become more distorted. From this information, we can see that the population is aging at a rate of 26.0% and therefore, Japan is a super-aged society.

In the "World Population Prospects: The 2012 Revision," published by the United Nations, Japan ranked first in the percentage of people of 60 years of age or above at 33%, and Italy and Germany placed second at 27% [19].

In 2012, the medical care expenditure was 39.2 trillion yen and had increased by 627 billion yen compared with the previous year. This trend of rising medical care expenditure is not specific to the years under consideration. Additionally, 15% of this medical care expenditure was for lifestyle-related health care costs. This represents a significant cost burden.

The main diseases that afflict the population of Japan have changed from being infectious diseases to being lifestyle-related diseases. The cause-specific mortality of lifestyle-related diseases is 60%, and this proportion may rise. The Ministry of Health, Labour, and Welfare stated that lifestyle-related diseases are not only the biggest factors in reducing healthy life expectancy, but also have the most significant impact on national medical care expenditure [11]. Lifestyle-related diseases can be prevented by moderate daily exercise, a well-balanced diet, and not smoking.

Morita et al. and Sasaki have reported the relationship BMI and Japanese diet [9,16]. In their paper, Western diet came to Japan in about 1955, and Japanese began to eat excess calories. In this paper, we discuss the relation between health statistics and eating habits, such as the number of medical checkup examinees and diet by prefecture using Annual Health, Labour, and Welfare report, 2008 of the Ministry of Health, Labour, and Welfare [26]. In addition to this, we discuss the relation between the number of patients and factors related to lifestyle such as body mass index and foods intake by prefecture [27].

We use a fuzzy regression model for the analyses. Generally, the relation between medical care expenditure and diet varies widely, and it is difficult to accurately describe the features of a target system. Therefore, a fuzzy robust regression model developed by Yabuuchi and Watada [25] is used to analyze the relation between medical care expenditure and factors related to lifestyle. This is because this model can illustrate the possibility of a target system even if the data used is irregular or the features of the system are not clear.

* Corresponding author.
E-mail addresses: yabuuchi@shimonoseki-cu.ac.jp (Y. Yabuuchi), kawaurat@hirakata.kmu.ac.jp (T. Kawaura), junzow@osb.att.ne.jp (J. Watada).

http://dx.doi.org/10.1016/j.compbiomed.2016.01.010
0010-4825/

As mentioned above, we analyze data for two indicators by prefecture in order to discuss the relation between the health indicators and lifestyle using a fuzzy robust regression model. The first analysis is that of the relation between the medical care expenditure and the factors indicating lifestyle. The second analysis is that of the relation between the number of patients and the factors indicating lifestyle.

This paper is organized as follows. In Section 2, a fuzzy robust regression model is introduced. A fuzzy robust regression model is able to remove the influence of unusual samples and to illustrate the possibility of a focal system. Then we employ a fuzzy robust regression model to analyze medical care expenditures and the number of outpatients by factors related lifestyle. In Section 3, we discuss the relation between medical care expenditure and factors of people's personal life, such as the number of medical checkup examinees and diet by prefecture. In Section 4, we discuss the relation between the number of outpatients and lifestyle factors such as a body mass index and food intakes. Finally, this paper is concluded in Section 5.

2. Fuzzy robust regression model

A fuzzy regression model can be categorized into two types. The first type is an interval model based on the possibility concept, and the second type is a non-interval model based on the least squares method. An interval model has been proposed by Tanaka and Guo [17], and a non-interval model has been proposed by Diamond [1,2]. The fuzzy robust regression model used in this study is based on the interval model proposed by Tanaka and Guo.

Since, observed data should embody the possibilities of the system under consideration, observed data can be interpreted as the possibilities of the system [3–6,17,20]. A fuzzy regression model is built in terms of possibilities and evaluates all observed values as possibilities that the system should contain. An interval fuzzy regression model is advantageous for an analysis of data with an ambiguity of person. In addition, an interval value is easy to understand a state of a system. Then we employ an interval fuzzy regression model in this paper.

Data sets $(y_i, \mathbf{x}_i)$ with an objective variable y_i and p explanatory variables $\mathbf{x}_i$ for $i=1,2,\ldots,n$ are given. A triangular shaped fuzzy regression coefficient $\mathbf{A}=(\mathbf{a},\mathbf{c})$ with a center $\mathbf{a}$ and a width $\mathbf{c}$ are used. Then, the fuzzy regression model can be defined as the following:

$$\mathbf{Y}_i=\mathbf{A}\mathbf{x}_i=(\mathbf{a},\mathbf{c})\mathbf{x}_i. \tag{1}$$

Here $\mathbf{Y}_i=(Y_i^C, Y_i^L, Y_i^U)$ is denoted with a triangular shaped fuzzy number with center Y_i^C, lower limit Y_i^L, and upper limit Y_i^U for $i=1,2,\ldots,n$.

The inclusion relation $Y_i^L \le y_i \le Y_i^U$ between the model $\mathbf{Y}_i$ and the objective variable y_i is rewritten as follows:

$$\mathbf{a}\mathbf{x}_i-\mathbf{c}|\mathbf{x}|_i \le y_i \le \mathbf{a}\mathbf{x}_i+\mathbf{c}|\mathbf{x}|_i. \tag{2}$$

Here $|\mathbf{x}|_i=[|x|_{i1}, |x|_{i2}, \ldots, |x|_{jp}]$.

This fuzzy regression model is built to contain all the data in the model. When the width of the model is large, the expression of its regression equation is vague. Therefore, it is better and more convenient to obtain a clear and rigid expression. Hence, the width of the regression should be minimized in order to reduce the vagueness of the model as much as is possible. The fuzzy regression model is formulated to minimize its width under the constraints (2). This problem results in the following linear programming (LP):

$$\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{n} \mathbf{c}|\mathbf{x}_i| \\
\text{subject to} \quad & \mathbf{a}\mathbf{x}_i+\mathbf{c}|\mathbf{x}|_i \ge \mathbf{y}_i \\
& \mathbf{a}\mathbf{x}_i-\mathbf{c}|\mathbf{x}|_i \le \mathbf{y}_i \\
& \mathbf{c} \ge \mathbf{0} \ (i=1,2,\ldots,n).
\end{aligned} \tag{3}$$

Solving the above LP problem gives a possibilistic regression. This fuzzy regression contains all data in its width and therefore the results express all possibilities that data can embody and the system under consideration should have.

The pivotal role of the center position of the system is emphasized when building a possibilistic regression model instead of using an interval to describe the possibility of a focal system. Some interval fuzzy regression models have been proposed to make the centers of a possibility distribution and the center of a possibilistic regression model coincide [5,17,22–24].

The model by Yabuuchi and Watada [22] describes the possibility of a system using the center of a fuzzy regression model. The proposed model fits intuitive understanding because it makes the center of the model and the center of the system coincide.

This fuzzy robust regression model is built by maximizing the sum of possibility grades derived from model estimates and data. In other words, the model is built to illustrate the possibility distribution. Therefore, our model can be built using granular or fuzzy data, as the model can deal with both granular and fuzzy data.

Let us calculate the total sum Z_1 of possibility grades of the fuzzy regression model as follows:

$$Z_1=\sum_{i=1}^{n}\mu(y_i,\mathbf{x}_i),\ \mu(y_i,\mathbf{x}_i)=\begin{cases}\max\left(0,1-\dfrac{|Y_i^C-y_i|}{W_i^L}\right); & y_i<Y_i^C\\[2ex] \max\left(0,1-\dfrac{|Y_i^C-y_i|}{W_i^U}\right); & Y_i^C<y_i,\end{cases} \tag{4}$$

Here W_i^L and W_i^U are the lower and upper widths of the model $\mathbf{Y}_i$, respectively. The model is built by maximizing Z_1, as defined in Eq. (4), which is the total sum of possibility grades in the objective function of a fuzzy regression model. It is clear from Eq. (4) that the width W_i of the model becomes larger if Z_1 is maximized. Therefore, the following function is defined to minimize the sum Z_2 of the vagueness values, W_i, of the model:

$$Z_2=\sum_{i=1}^{n} W_i. \tag{5}$$

Thus, the model is reduced to a bi-objective LP problem employing Z_1 and Z_2 as a multi-objective function.

However, it is easy to build a method to solve the problem using the following weighted sum of the two objective functions:

$$\begin{aligned}
\text{maximize} \quad & Z_3=\alpha Z_1-(1-\alpha)Z_2 \\
\text{subject to} \quad & Y_i^L \le y_i \le Y_i^U \quad (i=1,2,\ldots,n)
\end{aligned} \tag{6}$$

Here α is a weight parameter, $0 \le \alpha \le 1$. It is possible to control the shape of the model by changing the parameter α of the objective function. Therefore, the value of the parameter α is selected heuristically and empirically by the decision-maker.

Outliers distort the possibilities described by an interval fuzzy regression model, such as the model by Tanaka and Guo and our model. Controlling for the effects of outliers is very important in order to illustrate the possibility that the system has by an interval fuzzy regression model.

When data is widely distributed, it is hard to make the centers of the model and data distribution coincide. Eq. (4) shows that a possibility grade μ can be obtained using data $(y_i, \mathbf{x}_i)$ with a fuzzy regression model $\mathbf{Y}_i$. A greater data width distorts the shape of the model, and the center of the model no longer coincides with that of data distribution. This problem can be solved by enlarging the

Please cite this article as: Y. Yabuuchi, et al., Analysis of the relation between health statistics and eating habits in Japanese prefectures using fuzzy robust regression model, Comput. Biol. Med. (2016), http://dx.doi.org/10.1016/j.compbiomed.2016.01.010