International Journal of Business and Management
www.ccsenet.org/journal.html
Vol. 4, No. 10
October 2009

Evaluation of Beijing Urban Residents' Resource Worrying Consciousness Based on the Logistic Regression Model

Jiliu Sun
School of Statistics, Capital University of Economics and Business
Beijing 100070, China
Tel: 86-10-6760-7956 E-mail: sunjiliu1956@163.com

Abstract

According to the data in the survey of the urban residents' resource conservation index of Beijing in July of 2008, the binary choice model, i.e. the Logistic regression model, is used to compare and study residents (different sexes and different ages)' worrying degrees about the actuality of resources in China. The result of the quantitative research indicated that Chinese residents had strong resource worrying consciousness, and older residents more worried about the actuality of resources in China, and female worrying degree was higher than male's.

Keywords: Logistic regression model, Resource conservation index

The data used in the article all come from the survey of the urban residents' resource conservation index of Beijing in July of 2008. Because the research objects are residents (with different sexes and ages)' worrying degrees about the actuality of Chinese resources, so other contents in the questionnaires will not be analyzed in the article.

The survey result shows that 71.4% of respondents worried about the actuality of Chinese resources, and they thought Chinese reserves of oil and natural gas were limited, and the water resource and electric resource were deficient, and to continually keep the stable and quick development of macro economy, Chinese governments in all levels face serious challenges.

The attributive variable in the research *Q*, i.e. residents' worrying degree about the actuality of Chinese resources, is 0-1 variable, and the numerical value of 1 denotes worrying, and 0 denotes not worrying, and the independent variable, *Sex*, is denoted by 1 (male) and 2 (female), and the independent variable, *Age*, includes 9 classes, and 1 denotes "below the age 12 ", and 2 denotes "the age 12-15 ", and 3 denotes " the age 15-18", and 4 denotes "the age 18- 23 ", and 5 denotes "the age 23- 30", and 6 denotes "the age 30 - 40 ", and 7 denotes "the age 40 - 50 ", and 8 denotes "the age 50 - 60 ", and 9 denotes "the age 60 and above ", so the numerical value is bigger, the respondent's age is older. It is not feasible to use traditional statistical methods such as the variance analysis or the regression analysis, because they all require that the attributive variables are continuous, so other statistical analysis methods should be considered. The Logistic model can better solve this problem. On the one hand, it requires that the attributive variable is data with 0-1 type, which can overcome the limitation of traditional statistical methods. On the other hand, it is more persuasive than the contingency analysis by quantitatively studying the contribution degree of the independent variable to the attributive variable, so it is a kind of useful statistical method to process the data of questionnaire.

1. Introduction of fundamentals

First, to confirm whether the residents (with different sexes)' worrying degrees about the actuality of Chinese resources are different or not, the crossing frequency table of *Sex* and *Q* (seen in Table 1) is plotted.

Because Table 1 is the dimension of 2×2, so its Chi Square test should adopt the *Fisher* rigorous test and the value of P produced by the *Fisher* rigorous test with the language of R is $0.07866<0.10$, and under the confidence level of 90%, there are sufficient reasons to reject the original hypothesis, i.e. different sexual residents' worrying degrees are significantly different.

Second, to confirm whether the residents (with different ages)' worrying degrees about the actuality of Chinese resources are different or not, the box-shape figures about the age under two kinds of worrying degrees are plotted (seen in Figure 1).

From Figure 1, the median of the age box-shape figure under the worrying condition is obviously higher than the median of the age box-shape figure under the condition of not worrying, which indicates that the distributions of age under two kinds of conditions are different.

In above analysis, the residents' worrying degrees about the actuality of Chinese resources are influenced by both sex and age, and the worrying degree belongs to the data of 0-1 type, so the Logistic regression model is adopted to research this problem.

2. Logistic regression model

Usual econometric models all suppose that the attributive variables are continuous, but in real economic decision-makings, many problems about choice are faced. People need to choose in selective limited projects, which is usually opposite with the hypothesis that the explained variable is continuous variable, so the attributive variable only adopts limited discrete values. For example, people's choices about vehicles usually include subway, bus and taxi, and the investment decisions include the stock and real estate. The econometric models which use these decision-making results as the explained variables are called the discrete choice model. In the discrete choice model, the simplest situation is to choose one project in two selective projects, and here the explained variables only include two values, and the model is called as the binary choice model.

In the binary choice model, if the simple linear regression equation is used to fit the 0-1 variable y, the phenomenon of heteroskedasticity will be produced in the residual error sequence of the model, and the fitted value of the model will not be between 0 and 1. Therefore, to solve this problem, the immeasurable hidden variable y^* is introduced to replace the attributive variable y to establish the regression equation with the independent variable.

$$y_i^* = x_i\beta + u_i$$

Where, u_i is the random disturbance term. Then define the relationship expression of the hidden variable y^* and the 0-1 attributive variable y.

$$y_i = \begin{cases} 1 & y_i^* > 0 \\ 0 & y_i^* < 0 \end{cases}$$

So, $E(y_i \mid x_i, \beta) = P(y_i = 1 \mid x_i, \beta) = P(y_i^* > 0) = P(x_i\beta + u_i > 0) = 1 - F(-x_i\beta)$, and F denotes the distribution function of u_i which is required as a continuously monotonic increasing function. Therefore, the original regression model can be regarded as the following regression model of y_i about its conditional mean.

$y_i = 1 - F(-x_i\beta) + u_i$

According to different F in the distribution function, the binary choice model generally includes the Probit model, the Logistic Model and the Extreme model.

For the parameter estimation of the model, the binary choice model generally adopts the maximum likelihood estimation, and the likelihood function is

$$L = \prod_{y_i=0}[1 - F(x_i\beta)]\prod_{y_i=1}F(x_i\beta)$$

The first-order condition of the logarithm likelihood function is

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{N}\left[\frac{y_i f_i}{F_i} + (1-y_i)\frac{-f_i}{(1-F_i)}\right]x_i = 0$$

Where, f_i denotes the probability density function of u_i. If the expressions and sample values of the distribution function and the density function of u_i are known, by solving the equation group, the maximum likelihood estimation value of the parameter of the binary choice model can be obtained.

The Logistic regression model is the binary choice model which supposes that the distribution function corresponded with u_i fulfills the logic distribution, i.e. $F = e^x / (1 + e^x)$, so the above method can be used to estimate the Logistic regression model, and the form of the equation is

$$\ln(\frac{p}{1-p}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p \quad \text{or} \quad p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p)}}.$$

Where, p denotes the probability when the attributive variable y is 1. Here, the coefficient β of the model estimation can not be explained as the marginal influence to the attributive variable y, and it can be judged from the symbols only, i.e. if the coefficient is positive, it denotes that if the explained variable is bigger, the probability that the attributive variable is 1 is bigger, and contrarily, if the coefficient is negative, it denotes that the corresponding probability is smaller.

3. Data analysis

The *lrm* order in the language R of the Design software package is applied to implement the Logistic regression analysis for the data, and the regression result is

$$\ln(\frac{p}{1-p}) = 0.2175 + 0.0998Age + 0.2446(Sex = 2)$$

$$z = (0.78) \quad (2.12) \quad (1.67)$$

$R^2 = 0.012$ $\quad$ $L.R. = 7.68$

From the Logistic regression result, under the confidence level of 90%, the coefficients of the model all pass the Z test, and both the fitting degree R^2 and the total significance test statistics *L.R.* (the maximum likelihood rate) are big, so the fitting effect of the Logistic regression model is good, and it can be used to analyzed and evaluated.

First, the residents (with different ages)' worrying degrees about the actuality of Chinese resources will be analyzed. The coefficient of the age in the Logistic regression model is 0.0998, and it indicates that when the sex is certain and the age is enhanced each unit, the probability *P* that the respondent worries about the actuality of resources will be higher 0.0998 unit than the probability *1-P* that the respondent doesn't worries about the actuality of resources. In another words, the age is older, the probability worrying about the actuality of resources *P* is larger.

In the same way, the residents (with different sexes)' worrying degrees about the actuality of Chinese resources will be analyzed. Because the variable of the sex is introduced into the model as the dummy variable, so the model can be regarded as two models, i.e. the Logistic regression model of the age to the worrying degree under the condition of males and the Logistic regression model of the age to the worrying degree under the condition of females. The concrete model is seen as follows.

The model of males: $\ln(\frac{p}{1-p}) = 0.2175 + 0.0998Age$

The model of females: $\ln(\frac{p}{1-p}) = 0.4622 + 0.0998Age$

In two models, under condition of same age, the female advantage is obviously higher than males, which indicates that the female worrying consciousness about resources is higher than male worrying consciousness about resources.

Thus, the logistic models with two levels about 0-1 in the evaluation of residents' worrying degrees about resources were established, and the influences of different sexes and different ages on the worrying degrees of resources were compared in the article. The analysis result indicates that different sexes and different ages will largely influence residents' worrying degrees about resources, and their influences are all positive, and females' worrying consciousness is stronger than males' worrying consciousness, and the age is older, the residents' worrying consciousnesses about resources are stronger. From the estimation and test results of the model, the application of the Logistic regression model in the evaluation of the residents' worrying consciousness about resources is successful.

References

Gao, Tiemei. (2006). *Econometric Approach and Modeling.* Beijing: Tsinghua University Press, P. 200-204.

Table 1. Crossing frequencies of sexes and worrying degrees

	Not worrying 0	Worrying 1
Male 1	129	278
Female 2	135	377

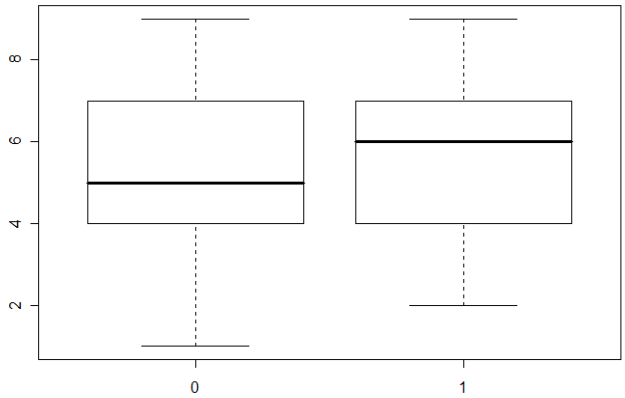

Figure 1. Box-shape Figure of Age and Worrying Degrees