Available online www.jsaer.com

Journal of Scientific and Engineering Research, 2020, 7(6):155-159

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Traffic Mode Selection Prediction Model Based on Standard Multinomial Logit Model

Zhenggan Cai[1]*, Shangshang Hu[1], Zhen Zhang[2], Jie Guo[3], Weizhi Li[4], Xiaohao Guo[5]

[1]School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo, Shandong, China
[2]School of Mining, China University of Mining and Technology, Xuzhou, Jiangsu, China
[3]CCCC-SHB Sixth Engineering Co., Ltd., Xian, Shanxi, China
[4]School of Science, Nanjing University of Science and Technology, Nanjing, Jiangsu, China
[5]Shouguang Office of Shandong Fishery Mutual Insurance Association, Shouguang, Shandong, China
*Corresponding author: zhenggancai@163.com

Abstract To further quantitatively assess the impact of different factors on the choice of residents' travel modes. By investigating the travel of residents in Hefei, from the perspective of transportation planning, a Standard Multinomial Logit (MNL) transportation mode prediction model was built. And the maximum likelihood method was used to measure the model parameters. The results show that: (1) the goodness of fit of the model and the Independence of irrelevant alternatives (IIA) hypothesis test meet the requirements. (2) Sex, private cars, driving licenses, peak hours, travel purposes, income, age have a significant impact on the choice of residents' travel mode. The research results can provide theoretical reference for traffic management departments.

Keywords Traffic engineering, Traffic planning, Multinomial Logit Model, Traffic planning

Introduction

With the development of China's economy, there are more and more alternative modes of transportation, which brings severe challenges to urban traffic planning. As an important step in the process of traffic planning, traffic mode prediction plays an indispensable role in urban planning. It sees the differences between different types of transportation. Thus, in the initial planning process, we can consciously deal with this difference to maximize the efficiency of municipal operation.

At present, many scholars at home and abroad have studied the influencing factors of residents' choice of travel mode, and found that age, travel distance, residence infrastructure and other factors have a greater impact on Residents' choice of travel mode [1-3]. For example, when the natural environment around the residence is good, travelers are more likely to travel for a short distance, and prefer slow-moving transportation such as bicycle [4]. Compared with the residents living in the city center, the travelers living in the outskirts of the city are more inclined to choose the car as the travel mode, and the residents in different regions have different choices of the traffic mode when they travel [5]. Residents near the rail transit station are 5-7 times more likely to choose rail transit travel mode than those far away from the rail transit station [6]. In addition, in the process of choosing traffic mode, travelers are also strongly influenced by their travel purpose. For instance, in urban areas, travelers usually prefer to choose fast means of transportation for business purposes. If the purpose of travelers is to play, they usually choose slow traffic instead of gaining traffic. Therefore, the travel of urban residents has obvious regularity. Corresponding to the city, the economic development in rural areas is relatively weak, and the time concept of travelers is generally lower than that in urban areas. The travel is greatly affected by the external environment, and the regularity is also lower than that in urban areas. Therefore, the travel behavior of urban residents has been widely concerned by scholars [7-8].

Nowadays, the traffic problem has become a worldwide problem, especially in the rush hour. The large-scale traffic flow increases the complexity of traffic planning and design. This kind of problem is especially noticeable in the first tier cities in China. Therefore, from the perspective of traffic planning, the establishment of traffic mode prediction model and the study of the influencing factors of residents' choice of various traffic modes can provide a theoretical basis for traffic planners and traffic management departments.

Multinomial Logit Model (MNL)

Considering that when residents travel, there are many alternative modes of transportation, which is typically multi-category variables. Using logit model based on stochastic utility maximization theory is the most commonly used analytical method in the study of travel behavior selection [9-10]. Among them, the Standard Multinomial Logit Model (MNL) is the most widely utilized. Therefore, this paper uses the MNL model to establish a traffic mode selection model, and analyzes the impact of different factors on Residents' choice of travel mode.

Using the random utility method, suppose that the random utility that the *i*-th traveler can bring when selecting mode *j* is:

$$U_{ij} = x'_{ij}\beta_j + \varepsilon_{ij} \tag{1}$$

In the formula, U_{ij} is the effect when the *i*-th traveler chooses mode *j*; x_{ij} is the set of independent variables; ε_{ij} is the model intercept; β_j is the regression coefficient.

Assuming that ε_{ij} is independently identically distribution(IID) and obey the type I extreme value distribution, then it can be proved that the probability of the *i*-th traveler to choose mode *j* is:

$$P(y_i = j|x_i) = \frac{\exp(x'_i\beta_j)}{\sum_{j=1}^{J}\exp(x'_i\beta_j)} \tag{2}$$

When building MNL model, we usually take a scheme (*j*=1 in this paper) as the base category, and then make its corresponding coefficient $\beta_1 = 0$. The probability of the *i*-th traveler choosing mode *j* can be further expressed as:

$$P(y_i = j|x_i) = \begin{cases} \dfrac{1}{1+\sum_{j=2}^{J}\exp(x'_i\beta_j)} (j=1) \\ \dfrac{\exp(x'_i\beta_j)}{1+\sum_{j=2}^{J}\exp(x'_i\beta_j)} (j=2,...,J) \end{cases} \tag{3}$$

Data Collection

Considering that at this stage, the statistical mechanism of traffic travel is not perfect and relevant statistical data cannot be obtained directly. Based on the comprehensive consideration of the characteristics of individual and family attributes and travel characteristics of travelers, we combined the survey data of Hefei residents' travel to model. After systematic screening of the original data, a total of 1432 trip data were used for modeling and analysis. The variable names, discretization results and corresponding frequencies in this paper are given in Table 1.

Table 1: Data Statistics and Coding

Variable	Code	Frequency	Variable	Code	Frequency
Travel mode	Slow traffic mode(1)	947	Travel purpose	Home(0)	644
	Private car(2)	109		Company(1)	335
	Public tools(3)	376		School(2)	107
Sex	Male(0)	697		Shopping(3)	346
	Female(1)	735	Revenue(thousand)	⩾150(0)	228
private cars	No(0)	519		50-150(1)	1034
	Yes(1)	913		⩽50(2)	170
Driving license	No(0)	971	Mood	Happy(0)	1024
	Yes(1)	461		Others(1)	408
Education level	University (0)	308	Year	20-50(0)	894
	Others(1)	1124		⩾50(1)	367
Peak hours	No(0)	868		⩽20(2)	171
	Yes(1)	564	-	-	-

Model Analysis

Parameter estimation

The model takes the mode of travel as the dependent variable, and chooses 9 factors as the independent variable from the three aspects of traveler's personal characteristics and family characteristics. The parameters of the model are estimated by the maximum likelihood method, and the modeling process is implemented by stata16.0 software. Through the *p*-value evaluation of the significance of the parameters, if $p \leq 0.05$, it indicates that the parameters have a significant impact on the travel mode, if $p \leq 0.01$, it indicates that the parameters have a very significant impact on the choice of the traffic mode, and the non-significant factors are eliminated, and the final model is shown in Table 2.

Table 2: model parameter estimation

Variables	Private car			Public tools		
	Coef.	S.E.	RR	Coef.	S.E.	RR
Female	0.694**	0.024	1.94	-0.135*	0.051	0.61
Have private car	0.350**	0.017	1.41	-	-	-
Driving license	0.509**	0.018	1.66	0.274*	0.006	1.32
Peak hours	-0.749**	0.025	0.38	0.871**	0.039	2.39
Company	-	-	-	0.479**	0.028	1.61
School	-	-	-	0.264**	0.018	1.32
Shopping	0.138**	0.071	1.14	-	-	-
Revenue≤50	-	-	-	2.150**	0.075	5.74
Year⩾50	-	-	-	-0.782**	0.096	0.46
Year⩽20	-	-	-	1.369*	0.044	3.46
Intercept	-0.530	0.098		-1.516	0.043	
Log Likelihood	-1684.509					
McFadden R^2	0.307					

Note: *for $p \leqslant 0.05$; **for $p \leqslant 0.01$.

It can be seen from table 2 that 10 variables such as female travelers, private cars, driver license, peak hours, company, school, shopping, annual income of 50-150, annual income ≤ 50, age ≥ 50, age ≤ 20 have statistically significant effects on travelers' choice of transportation mode.

a. Female travelers are 1.94 times more likely to choose private cars as their mode of travel than slow-moving vehicles, and 0.61 times more likely to choose public vehicles; b. Travelers with private cars are 1.41 times more likely to choose to drive private cars as their way of travel than those who choose chronic means of transportation; c. The probability of driving private cars and public transportation for licensed residents is 1.66 times and 1.32 times higher than that for slow-moving vehicles; d. In peak hours, the probability that residents choose to drive private cars is lower than that of choosing slow-moving vehicles, while the probability of choosing public vehicles is 2.39 times higher than that of choosing slow-moving vehicles. This may be because the traffic flow in peak hours is large, and it is easy to get stuck when driving. Therefore, residents who travel in peak hours generally do not Choose the way of driving private cars, while most residents will choose the subway, bus and other public transport as the way of travel; e. When the purpose of residents' travel is to go to work, the impact of private cars on the choice of transportation mode is not statistically significant, while the probability of choosing public transportation is 1.61 times that of choosing slow-moving vehicles; f. When the purpose of residents' travel is school, the probability of choosing public transport is 1.32 times of choosing slow-moving transport; g. For the purpose of shopping, the probability of choosing private car as the travel mode is 1.14 times of choosing slow traffic mode; h. Compared with the residents whose annual income is higher than 150000, the probability of choosing public transport for the residents whose annual income is lower than 50000 is increased by 5.74 times; i. Compared with travelers aged 20-50, those aged over 50 have a lower probability of choosing public transport than those aged over 50; j. Residents younger than 20 years old are 3.46 times more likely to choose public transport.

Goodness of fit test

After the estimation of model parameters, in order to determine the fitting effect of the model on the data, it is necessary to test the goodness of fit of the model, using Pearson statistics and deviation statistical indicators to test the goodness of fit of the model. The specific model test results are shown in Table 3, Chi-square test is not significant, indicating that the model has a better fitting effect on the data.

Table 3: Goodness of Fit Test

Parameters	Chi-square test	Significance
Pearson	2091.79	0.786
Deviation	934.25	1.000

McFadden R^2 is another index of fitting test for the MNL model. The value range of the fitting index is 0-1. The closer it is to 1, the better the fitting effect of the model. The value of McFadden R^2 in this paper is 0.307, and the value of McFadden R^2 in reference [11] is 0.2-0.4. The fitting effect of the model can meet the requirements. The fitting value of the model built in this paper is in this range, so it can be seen that the fitting effect of the model is very good.

The test for Independence from Irrelevant Alternatives (IIA)

As the premise hypothesis of the establishment of MNL model is to meet the IIA hypothesis, it is necessary to conduct the independent hypothesis test on the model, and the test results are shown in Table 4. The test results show that omitting either of the private Car and Public Tools does not reject the IIA hypothesis. Moreover, the test results of slow traffic mode and private car as reference categories will not reject the IIA hypothesis, indicating that the IIA test has passed.

Table 4: Hausman tests of IIA assumption

Travel mode	$\chi 2$-Value	p-Value	Null hypothesis	IIA property
Slow traffic mode	-0.201	1.00	Fail to reject	Holds
Private car	0.000	1.00	Fail to reject	Holds
Public tools	0.000	1.00	Fail to reject	Holds

Conclusion

(1) The division and prediction of traffic mode is an important research content of traffic planning, so many traffic planning researchers study the related directions. From the perspective of traffic planning, based on the travel survey of Hefei residents, this paper collected 1432 sample cases. Taking the choice of traffic mode as the dependent variable, nine factors are selected as the independent variable from three aspects of traveler's personal characteristics, family characteristics and other characteristics, and a standard multiple logit model is

constructed. The results show that 7 factors, such as female driver, private car, driver's license, peak hour, travel purpose, income and age, have a significant impact on the choice of residents' travel mode. The research results can provide theoretical reference for traffic planning department and management department.
(2) In the follow-up research process, we can consider to further optimize the standard multinomial logit model, collect more comprehensive travel data, so as to build a more perfect traffic mode prediction model.

Reference

[1]. Sasidhar, K., Vineeth, Y., Vineethreddy, & Subbarao, S. S. V. (2016). Trip Attraction Rates of Commercial Land Use: A Case Study. *Indian Journal of Science & Technology*, 9(30), 1-5.
[2]. Razib, S. A., Rahman, F. I. (2017). Determination of trip attraction rates of shopping centers in Uttara Area, Dhaka. *American Journal of Management Science and Engineering*, 2(5), 150.
[3]. Meena, S., Patil, G. R., Mondal, A. (2019). Understanding mode choice decisions for shopping mall trips in metro cities of developing countries. *Transportation Research Part F: Psychology and Behaviour*, 64, 133-146.
[4]. Friedman, B., Gordon, S. P., & Peers, J. B. (1994). Effect of neo-traditional neighborhood design on travel characteristics. *Transportation Research Record*, 1466, 63-70.
[5]. Bagley, M. N., Mokhtarian, P. L. (2002). The impact of residential neighborhood type on travel behavior: A structural equations modeling approach. *The Annals of Regional Science*, 36(2), 279-297.
[6]. Cervero, R. (1993). Ridership impacts of transit focused development in California. Berkeley, CA, USA: University of California.
[7]. Yang, L., Zheng, G., Zhu, X. (2013). Cross-nested logit model for the joint choice of residential location, travel mode, and departure time. *Habitat International*, 38, 157-166.
[8]. Yang, J., Ma, J. (2019). Compressive sensing-enhanced feature selection and its application in travel mode choice prediction. *Applied Soft Computing*, 75, 537-547.
[9]. Jing, P., Juan, Z. C., Zha, Q. F. (2014). Incorporating psychological latent variables into travel mode choice model. *China Journal of Highway and Transport*, 27(11), 84-92. (in Chinese)
[10]. Zhou, W., Zhao, S. (2013). Quantitative analysis of traveler route choice behavior based mixed Logit model. *Journal of Jilin University (Engineering and Technology Edition)*, 43(2), 304-309. (in Chinese)
[11]. Mujalli, R. O., De-Ona, J. (2013). Injury severity models for motor vehicle accidents: a review. *Proceedings of The Institution of Civil Engineers-transport*, 166(5), 255-270.