Available online www.jsaer.com

Journal of Scientific and Engineering Research, 2022, 9(8):59-67

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Research on Travel Mode Selection Decision Method Based on Discrete Choice Model

Yingcui Du, Benxing Liu, Shulong Dong, Xiaoqing Wang, Pengsheng Zhao

School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo 255000, China

Abstract With the development of Sweden's economy, new modes of transportation have gradually developed. In order to understand the acceptance of new modes of transportation by the Swedish people, this paper takes the Swedish people as the research object. The data are obtained by filling out questionnaires on the spot, mailing questionnaires and SP surveys. BIOGEME is used to establish a Multinomial Logit (MNL) model to analyze the travel choice behavior of the Swedish people under various modes of transportation, helping the government to further optimize and improve transportation operations and providing more efficient services to the people.

Keywords discrete selection model; MNL model; BIOGEME

1. Introduction

1.1 Research Status at Home and Abroad

Traffic mode division and traffic mode selection model is one of the main research directions in the field of traffic planning. At present, there are mainly two kinds of research methods for traffic mode Division: one is the traditional aggregate method. However, because traffic mode selection involves multiple factors and the personal characteristics of travelers will also have an impact, the traditional aggregate model can't meet the description of the characteristics of travelers' travel mode selection. The second type is the dis aggregate method. The non-aggregate model can calibrate the parameters of the model with fewer samples; Factors related to personal decision-making can be selected as independent variables to more accurately describe the travel decision-making process of individuals or families; It has better time transfer and regional transfer, and the application scenarios are more extensive, making up for the limitations of the aggregation model.

Daniel McFadden studied and laid the theoretical foundation of discrete selection model [1], and established the foundation for the theoretical research of non-aggregate model. The non-aggregate discrete selection model can analyze the travel purpose, travel mode and route selection from the individual level, and describe the macro traffic characteristics. Among them, the most commonly used one is the multinomial logit model (MNL), which has the advantages of closed probability and convenient parameter estimation; It is derived from the assumption that the error term satisfies the independent uniform distribution, so it has the characteristic of independent - selection scheme.

In recent years, domestic scholars have gradually used the non-aggregation theory to study individual travel characteristics, and there are more systematic theoretical studies and applications under various traffic scenarios. In 2004, Professor Guan Hongzhi introduced in detail the calibration methods and tests of BL model, MNL model and NL model, as well as the application of the model, which laid a foundation for the application of non-aggregate model [2]. In 2007, Chen Tuansheng studied the commuting behavior of residents, established MNL model and NL model, combined with cross elasticity coefficient and direct elasticity coefficient to analyze and describe the departure time, travel chain and transportation mode selection. Chang Chaofan studied the individual travel characteristics of short trips [3], and established NL model to study the residents' short trip mode selection. Li Liang analyzed the utility and decision-making of individual travel mode selection, and built MNL model to study the choice of individual travel mode [4], and discussed the decision-making process and mode of individual travel. Based on the logit model [5], LV Xiangru selected Lanzhou transport corridor as an example, established a multiple logit model with passenger characteristics and travel characteristics as variables, studied the passenger flow sharing rate of transport modes, and reached conclusions and suggestions on improving public transport and opening intercity railway. Liu Zhen, Zhou Xizhao and others [6] studied nested logit based on discrete selection theory and its application in transportation mode selection behavior, and proposed a set of application methods that are convenient for practical operation, and illustrated the whole application process with a numerical example. In addition, Chen Tuansheng [7] established NL model to study the travel mode selection of commuters based on activities. Zhao Lin [8] established the route selection model of travelers based on the expected utility theory and prospect theory, and conducted a comparative study through theoretical calculation, empirical investigation and simulation analysis.

1.2 Theoretical Basis

Based on the random utility theory, the discrete choice model assumes that the travel mode selection set of a traveler n is A_n. The utility of selecting transportation mode j is U_{jn}, then the conditions for the traveler n selects i from the selection set A_n are:

$$U_{in} \geq U_{jn}, i \neq j, i.j \in A_n$$

According to the random utility theory, utility is a random variable, and the utility function u can be divided into two parts: fixed term and random term. It is assumed that there is a linear relationship between them. Therefore, the utility of the traveler U_{in} can be expressed as:

$$U_{in} = V_{in} + \varepsilon_{in}$$

U_{in}— true utility of traffic mode *i* for user n;

V_{in}——for user *n*, utility function determination term of traffic mode *i*;

ε_{in}——utility random (error) term of traffic mode *i* for user *n*.

According to the random utility theory, the probability P_{in} of traveler *n* choosing traffic mode *i* can be expressed as:

$$P_{in} = Prob(U_{in} > U_{jn}; i \neq j, j \in A_n) = Prob(V_{in} + \varepsilon_{in} > V_{jn} + \varepsilon_{jn}; i \neq j, j \in A_n)$$

In the formula: $0 \leq P_{in} \leq 1. \sum_{i \in A_n} P_{in} = 1$

The selected probability function has the following properties:

(1) The constant term is added to the utility function, and the selection probability is unchanged.

(2) Utility function expansion$\alpha(> 0)$, the selection probability remains unchanged.

Based on these two properties of the selection probability function, by assuming the distribution form of the probability term, a variety of corresponding non-aggregate models can be derived. The main models are logit model, generalized extreme value model, probit model and mixed logit model. Among them, the logit model makes the random term obey the independent extremum distribution, and has the characteristic that the non-related options are independent of each other.

Assumptions in this paper ε_{in} follows the double exponential distribution (Gumbel distribution), and the distribution function of Gumbel distribution is as follows:

$$F(x) = e^{-e^{-\omega(x-\eta)}}$$

$$f(x) = \omega e^{-\omega(x-\eta)} \cdot e^{-e^{-\omega(x-\eta)}}$$

Supposing the value of (η, ω) is (0,1). According to the properties of Gumbel distribution, the general formula of multinomial logit (ML) model is as follows:

$$P_{in} = \frac{e^{V_{in}}}{\sum_{j\epsilon A_n} e^{V_{jn}}}$$

P_{in}— For user n, the probability of selecting traffic mode i ($i = 1,2,... n$);

V_{in}— For user n, utility function determination term of traffic mode i;

A_n— For user n, the set of transportation mode selection schemes.

2 Data

This paper uses the data provided by biogeme's official website. This data was collected on the railway train. The respondents' choice of travel mode was collected in the form of questionnaires. A total of 470 people were interviewed, of which 441 were available. Each of the 441 respondents was given the choice of 9 statements, providing three choices: train, Swiss subway and car (only for people with cars).

This sample uses video to observe the license plate of the expressway and sends a questionnaire to the vehicle owners. As of April 1998, 9658 letters were sent and 1758 letters were returned. A total of 1070 people completed the complete survey and were willing to participate in the second SP survey, which was generated using the same method as the railway interview. From the license-based surveys, 750 available SP surveys were returned.

The variables of the data set are shown in Table 1 and table 2, and the descriptive statistics are shown in Table 4. Table 3 includes the codes of the States.

Table 1: Variables and descriptions

Variable	Variable description
GROUP	Different groups in the population. 2: Current railway users; 3: Current road user
SURVEY	It is equivalent to grouping but uses different codes: 0: train user, 1: car user
SP	Fixed to 1 (statement preference survey)
ID	User identification
PURPOSE	The purpose of the trip. 1: Commuter, 2: shopping, 3: business, 4: leisure, 5: return from work, 6: return from shopping, 7: return from work, 8: return from leisure, 9: other
FIRST	First class passengers (0 = no, 1 = yes)
TICKET	This is my ticket. 0: none, 1: two-way half price card, 2: one-way half price card, 3: two-way normal, 4: one-way normal, 5: half day, 6: annual season ticket, 7: junior or senior season ticket, 8: free ride card after 7 pm, 9: group ticket, 10: other
WHO	Who pays (0: unknown, 1: self, 2: employer, 3: half and half)
LUGGAGE	0: No, 1: one piece, 3: several pieces
AGE	The age coding scheme belongs to the following types: 1: age<=24; 2: 24<age<=39; 3: 39<age<=54; 4: 54<age<=65; 5: 65<age; 6: not known
MALE	Gender of tourists: 0: female, 1: Male
INCOME	Annual passenger income [thousand Swiss francs] 0 or 1: less than 50, 2: between 50 and 100, 3: greater than 100, 4: unknown
GA	The variables capture the impact of the Swiss railway system and the annual

	season tickets for most local public transport. 1 if the individual owns GA, otherwise 0.
ORIGIN	Travel source (corresponding to the number of a state, see Table 3)
DEST	Tourist destination (corresponding to the number of a state, see Table 3)
TRAIN_AV	Train availability
CAR_AV	Vehicle availability
SM_AV	SM availability
TRAIN_TT	Train travel time [minutes]. Travel time is based on the assumption of car based distance (1.25 * bus distance)
TRAIN_CO	Train cost (CHF). If the passenger has GA, this fee is equal to the annual ticket fee.
TRAIN_HE	Train departure interval [minutes] For example, if there are two trains per hour, the departure interval of trains is 30min.
SM _TT	SM travel time [minutes] considering the future, the speed of swiss metro is 500 km / h
SM _CO	SM cost [CHF] without considering GA, it is calculated according to the current relevant railway fare and multiplied by a fixed factor (1.2) to reflect higher speed.

Table 2: Variables and description

Variable	Description
SM_HE	SM departure interval [minutes]
SM _SEATS	Seats are arranged in swiss metro (virtual). Flight seat (1) or (0).
CAR_TT	Car travel time [minutes]
CAR_CO	The fixed average cost per kilometer (1.20 CHF / km) is considered for the vehicle cost
CHOICE	Select the indicator. 0: unknown, 1: train, 2: SM, 3: Car

Table 3: Coded States

Number	Canton	Number	Canton	Number	Canton
1	ZH	10	FR	19	AG
2	BE	11	SO	20	TH
3	LU	12	BS	21	TI
4	UR	13	BL	22	VD
5	SZ	14	SH	23	VS
6	OW	15	AR	24	NE
7	NW	16	AI	25	GE
8	GL	17	SG	26	JU
9	ZG	18		GR	

1) Basic personal characteristics

This part is mainly to clarify the composition of the survey objects and the differences between the survey individuals. Basic personal characteristics include age, gender, income, whether there is a season ticket, etc. Through these basic attributes, individual differences can be distinguished, and it is convenient to analyze the travel choices of travelers under different economic conditions and working conditions. See **Table 2** and **Table 3**.

Table 4: Basic identity of travelers

Gender	Male	Female
	0	1

Table 5: Basic characteristics of travelers

Disposable income	<50	Between 50 and 100	>100	none
	0 or 1	2	3	4
Vehicles owned	none	train	SM	automobile
	0	1	2	3

2) Travel attributes

This part includes travel time, travel purpose, travel mode, travel frequency, travel cost and time consumption, etc. it is mainly to understand the travel time of the transportation mode selected by the people, and analyze the relationship between them in combination with their travel purpose and travel time. Secondly, a survey was also conducted on the main factors considered in the choice of travel mode to understand the respondents' emphasis on the economy, convenience, comfort, safety and punctuality of the transportation mode, which will be used as a reference in the modeling.

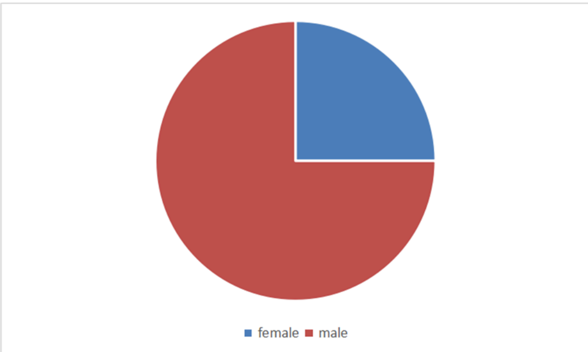

Figure 1: Gender distribution of respondents

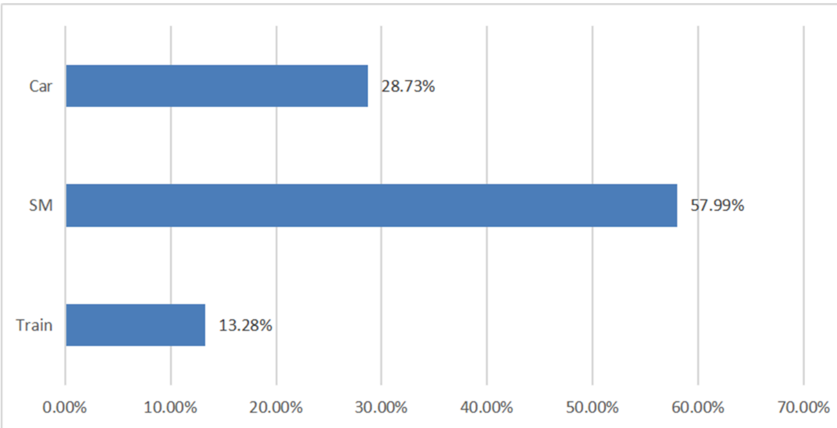

Figure 2: Distribution of travel mode selection

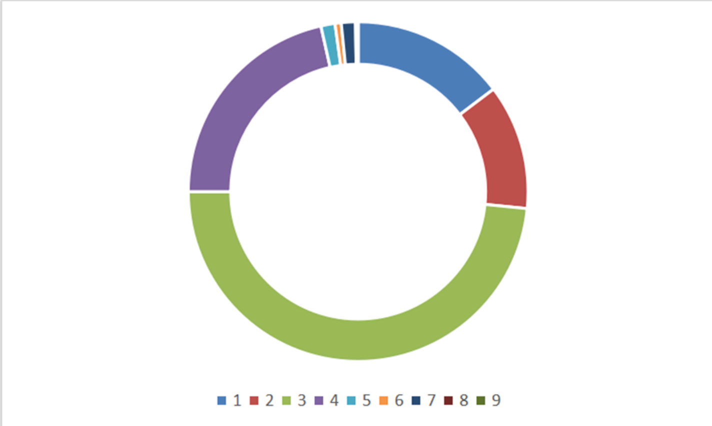

Figure 3: Travel destination distribution

3. Model Calibration

3.1 Data Preparation

Table 6: Definition of utility variables

Transportation	constant term	travel time	travel cost	travel time coefficient	travel cost coefficient
CAR	α_1	X_{1T}	X_{1C}		
SWISSMETRO	α_2	X_{2T}	X_{2C}	β_1	β_2
TRAIN	α_3	X_{3T}	X_{3C}		

$$X_{1T} = \frac{CAR_{TT}}{100} \quad X_{1C} = \frac{CAR_{CO}}{100}$$
$$X_{2T} = \frac{SM_{TT}}{100} \quad X_{2C} = \frac{SM_{CO}}{100}$$
$$X_{3T} = \frac{TRAIN_{TT}}{100} \quad X_{3C} = \frac{TRAIN_{CO}}{100}$$

3.2 Parameter calibration results

Based on the stochastic utility maximization theory, travelers always choose the transportation mode with the greatest utility relative to themselves. By comparing the utility values of the selected branches, the probability formula can be used to calculate the selection probability of the traveler's traffic mode. The utility function of each transportation mode is as follows:

Table 7: Utility function of various modes of transportation

Transportation	CHOICE	Utility Function
CAR	1	$V_{1n} = \alpha_1 + \beta_1 \times X_{1T} + \beta_2 \times X_{1C}$
SWISSMETRO	2	$V_{2n} = \alpha_2 + \beta_2 \times X_{2T} + \beta_2 \times X_{2C}$
TRAIN	3	$V_{3n} = \alpha_3 + \beta_3 \times X_{3T} + \beta_3 \times X_{3C}$

Biogeme is used for parameter calibration in this paper. The parameter calibration results are shown in the figure:

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_CAR	-0.155	0.0432	-3.58	0.000349	0.0582	-2.66	0.00786
ASC_TRAIN	-0.701	0.0549	-12.8	0	0.0826	-8.49	0
B_COST	-1.08	0.0518	-20.9	0	0.0682	-15.9	0
B_TIME	-1.28	0.0569	-22.5	0	0.104	-12.3	0

Figure 2: Parameter calibration results

"Name" row in the table represents the name of the traffic mode and its selection and influencing factors, "Value" represents the corresponding parameter value, "std err" is the standard error, and "t-test" is the t-test value of the parameter value. In the statistical results, the standard error of the calibration parameters is less than 1, so the data fitting result is better.

In the results, the selection branch constant items were normal, which represented the average value of the influence of external variables on the selection behavior explained by the model. The travel cost and travel time parameters are negative, and the t-test value is large, indicating that the increase of time and cost will significantly reduce the probability of selecting the travel mode.

According to the calibration results, the probability function model of the traveler can be obtained by bringing them into the fixed term of the utility function:

$$P_{1n}=\frac{e^{V_{1n}}}{e^{V_{1n}}+e^{V_{2n}}+e^{V_{3n}}}=\frac{1}{1+e^{V_{2n}-V_{1n}}+e^{V_{3n}-V_{1n}}}$$

$$=\frac{1}{1+e^{-1.08\times X_{2T}-1.28\times X_{2C}+0.155+1.08\times X_{1T}+1.28\times X_{1C}}+e^{-0.701-1.08\times X_{3T}-1.28\times X_{3C}-0.155+1.08\times X_{1T}+1.28\times X_{1C}}}$$

$$P_{2n}=\frac{e^{V_{2n}}}{e^{V_{1n}}+e^{V_{2n}}+e^{V_{3n}}}=\frac{1}{1+e^{V_{1n}-V_{2n}}+e^{V_{3n}-V_{2n}}}$$

$$=\frac{1}{1+e^{-0.155-1.08\times X_{1T}-1.28\times X_{1C}+1.08\times X_{2T}+1.28\times X_{2C}}+e^{-0.701-1.08\times X_{3T}-1.28\times X_{3C}+1.08\times X_{2T}+1.28\times X_{2C}}}$$

$$P_{3n}=\frac{e^{V_{3n}}}{e^{V_{1n}}+e^{V_{2n}}+e^{V_{3n}}}=\frac{1}{1+e^{V_{1n}-V_{3n}}+e^{V_{2n}-V_{3n}}}$$

$$=\frac{1}{1+e^{-0.155-1.08\times X_{1T}-1.28\times X_{1C}+0.701+1.08\times X_{3T}+1.28\times X_{3C}}+e^{-1.08\times X_{2T}-1.28\times X_{2C}+0.701+1.08\times X_{3T}+1.28\times X_{3C}}}$$

3.3 probability of traffic mode selection

In 3.2, Biogeme is used for parameter calibration. According to the results of Biogeme calibration, the selection probability of car, swiss metro and train is calculated as follows:

Table 5: Selection probability of various modes of transportation

Mode of transportation	CAR	SWISSMETRO	TRAIN
Prediction probability	26.15%	60.43%	13.41%
Actual probability	28.73%	57.99%	13.28%
error	-8.98%	4.21%	0.98%

By comparison, it can be seen from **Table 5** that SWISSMETRO is the most likely mode of transportation to be used for commuting in Switzerland, accounting for about 60.43%, which is similar to the actual survey results. Comparing the predicted results with the actual survey results, it can be found that the difference between the predicted traffic mode selection probability and the actual situation is small, but the difference between the CAR predicted results and the actual situation is large. The comparison between the predicted results and the actual results is shown in **Figure 4**.

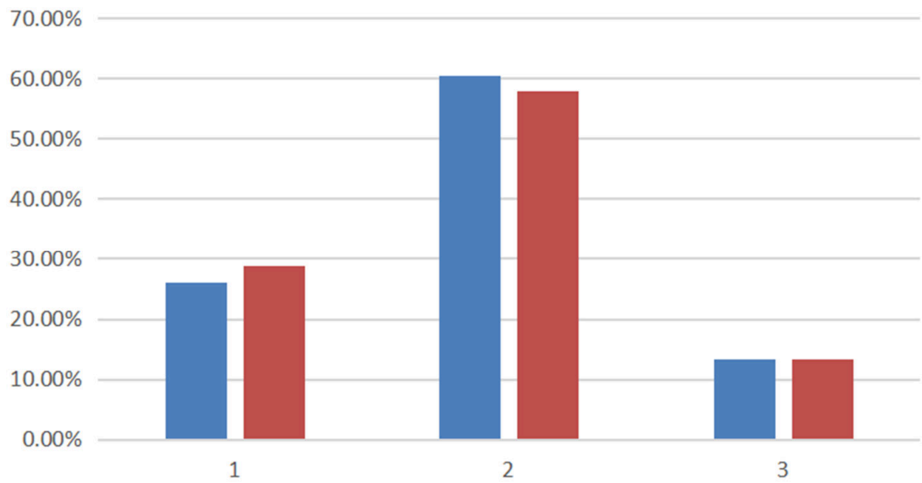

Figure 4: Comparison between prediction results and actual selection

The random utility theory assumes that all travelers can perceive the travel information in advance, and they are completely rational and can choose the travel mode with the highest utility. However, in actual life, travelers' perception range is limited, and they cannot obtain comprehensive traffic information. Moreover, travelers are not completely rational, and their personal psychology will affect their choice of travel mode, and they may not always adopt the most effective scheme. Therefore, there is a difference between the model prediction results and the actual survey results. In the future, the prospect theory can be considered to replace the random utility theory. The prospect theory can take the psychological and behavioral characteristics of travelers into account and make the prediction results more accurate.

4. Conclusion

In this paper, SP survey method is used to obtain the data of investigation on the travel choice of Swedish people, and MNL model based on random utility theory is established. Biogeme is used to calibrate the model parameters and conduct simulation test, determine the utility function of various travel modes, and carry out descriptive analysis on the selection of various travel modes. According to the research results, the surveyed Swedish people are mainly inclined to choose SWISSMETRO, a few people are inclined to choose CAR, and a very few people choose TRAIN. Finally, the reasons for the difference between the research results and the actual investigation results are analyzed, and the improvement scheme is proposed.

References

[1]. Mefadden D. Conditional logit analysis of qualitative choice behavior[M]. In Zarembka (ed.) Frontiers in Ecollometries. New York: Academic Press, 1973: 105-142

[2]. Guan Hongzhi. Non aggregate model -- a tool for traffic behavior analysis [M]. Beijing: People's Communications Press, 2004

[3]. Chang Chaofan. Research on short distance travel behavior of urban residents [D]. Beijing: Beijing Jiaotong University, 2007

[4]. Li Liang. Utility and decision analysis of urban residents' travel mode [D]. Wuhan: Huazhong University of science and technology, 2007

[5]. Lu Xiangru Study on passenger flow sharing rate of transport corridor based on logit model [J] Western communications science and technology, 201 5 (I): 82-87

[6]. Liu Zhen, Zhou Xizhao. Application of nested logit model in traffic mode selection [J]. Journal of Shanghai Maritime University, 2006 (3): 66-70

[7]. Chen tuansheng. Study on characteristics and analysis methods of commuter travel behavior [D] Beijing Jiaotong University, 2007

[8]. Zhao Lin Research on travel decision model and simulation experiment based on prospect theory Beijing: Beijing Jiaotong University, 2007