COMPARISON OF VARIOUS UTILITY FUNCTIONS FOR BEHAVIORAL TRAVEL DEMAND MODEL

by
Shigeru Morichi, Haruo Ishida* and Tetsuo Yai

Dept. of Civil Engineering
Tokyo Institute of Technology

2-12-1, O-okayama, Meguro-ku
Tokyo 152
Japan

1. Introduction

To date many different kinds of behavioral travel demand model have been developed. Of these the disaggregate Logit model has been utilized in an increasing number of cities for transportation planning.

Recently in Japan developers of large scale shopping centers and malls often have to confront with emotional opposition of local retailers who are afraid to lose their sales. This situation discourages the effort to improve the environment of shopping areas. To develop travel demand models for analysing the impact of shopping area renewal on shopping trip attraction and on the growth of income of local stores makes sense not only for transportation planning but also for urban planning.

Generally speaking, non-grocery shopping trips are under less constraints than home to work trip, so the way in deciding to make such trips might be more complicated than that for daily travel.

In this paper, a Logit model, a Fuzzy-integral model and a Lexicographic model are employed as destination-choice models and mode-choice models for non-grocery shopping trips and their performances are compared and evaluated.

2. The Principles for Travel Behavior and the Modelling Structure

In this paper the following three models are employed. The first one is a Logit model, that is the most frequently-used behavioral travel demand model. The second is a Fuzzy-integral model that has been developed to describe the vagueness and uncertainty of human thinking. The third model is a Lexicographic one which is quite different from the others, that is, the trip-maker is assumed to make decision by considering only one factor at a time.

2.1 Logit Model

The general formula of the logit model is as follows:

$$P(i:A) = \frac{e^{Vi}}{\sum_{j \in A} e^{Vj}} \qquad (1)$$

* Institute of Socio-Economic Planning, University of Tsukuba

and

$$Vi=\Sigma Xik\theta k \quad (2)$$

where

- $P(i:A)$: the probability that alternative i will be chosen
- A : choice set
- Vj : the systematic part of utility function
- Xik : value of each attribute
- θk : unknown parameters

2.2 Fuzzy-integral Model

The Fuzzy-set theory and Fuzzy-integral theory are employed to simulate the vagueness of someone's thinking[1]. A fuzzy set is characterized by a membership function which associates with each point in the fuzzy set a real number in the interval[0,1]. The nearer the vlaue of membership function to unity, the higher the grade of membership. On the other hand, for a set in the ordinary sense of the term, its membership function can take only two values 0 and 1.

In the Fuzzy-integral model, the membership function of each property which characterizes alternatives is estimated at first. The membership function, that is a kind of utility function for each factor, transfers the values of attributes(time, cost, etc.) to utility values. Then each utility value is combined together to determine the total utility. Suppose, there are only two attributes, x_{1i}, x_{2i}, which characterize an alternative i. The total utility of alternative i is given by eq.(3)[2][3].

$$U(i)=\{G_\lambda(1)\wedge\mu_1(x_{1i})\}\vee\{G_\lambda(2)\wedge\mu_2(x_{2i})\}\vee\{G_\lambda(1\cup 2)\wedge\mu_1(x_{1i})\wedge\mu_2(x_{2i})\} \quad (3)$$

where $U(i)$: total utility of alternative $i, V(i)\in[0,1]$
$\mu_j(x_{ji})$: membership function for attribute j, $\mu_j(x_{ji})\in[0,1]$
$G_\lambda(j)$: upper boundary of utility for attribute j
$G_\lambda(1\cup 2)=G_\lambda(1)+G_\lambda(2)+\lambda G_\lambda(1)\cdot G_\lambda(2)$
λ: parameter
and $\vee$ and $\wedge$ stand for max and min, respectively.

When there are n alternatives, the Fuzzy-integral model is as follows

$$U(i)=\underset{A'\subset A}{Max}[Min\{\underset{x_j\in A'}{Min}, [\mu_j(x_{ji})], G_\lambda(A')\}] \quad (4)$$

where A: universal space consists of attributes
A': arbitrary subspace of A
x_{ji}: value of attribute j for alternative i
$\mu_j(x_{ji})$: membership function for attribute j, $\mu_j(x_{ji})\in[0,1]$
$G_\lambda(A')$: upper boundary of utility for A'

$$G_\lambda(\phi)=0$$
$$G_\lambda(A)=1$$
$$G_\lambda(\underset{j}{\cup} X_j)=\frac{1}{\lambda}[\underset{j}{\pi}\{1+\lambda G_\lambda(x_j)\}-1]$$

λ: parameter

$G_\lambda(\cdot)$ is a kind of weight for each property. It has a similar role to the parameter θ in the Logit model. In eq.(3) and eq.(4), unknown parameters are $G_\lambda(\cdot)$ and λ. Parameters are estimated by maximizing the percent correct.

2.3 Lexicographic Model

For the Lexicographic model, an order of properties is determined at first from the survey data. Fig-1 schematically shows the decision-making process assumed for the Lexicographic model with two properties, time and cost[4][5][6]. If the most important factor is time, the choice probability for alternative 1 at the first stage is given by eq.(5).

$$P_1(A=1)=\alpha_1 \mathrm{Prob}(\Delta t>Ut) \tag{5}$$

where $\Delta t=|t_2-t_1|$

t_i: value of total travel time of alternative i (i=1,2)

$$\alpha_1=\begin{cases}1 & \text{when } t_2>t_1\\ 0 & \text{when } t_2\le t_1\end{cases}$$

Ut: value of threshold for total travel time

If the difference of total time between the two alternatives is smaller than Ut, then the alternatives are considered to be indifferent. When the two alternatives are indifferent concerning time, the probability that trip-makers will choose alternative 1 by considering cost is written as follows.

$$P_2(A=1)=\alpha_2 \mathrm{Prob}(\Delta c>Uc)(1-\mathrm{Prob}(\Delta t>Ut)) \tag{6}$$

where $\Delta c=|c_2-c_1|$

Ci: value of total cost of alternative i(i=1,2)

$$\alpha_2=\begin{cases}1 & \text{when } C_2>C_1\\ 0 & \text{when } C_2\le C_1\end{cases}$$

Uc: value of threshold for total cost

The probability that the two alternatives will be indifferent as for time and cost is given by eq.(7).

$$\overline{P}=(1-\mathrm{Prob}(\Delta t>Ut))(1-\mathrm{Prob}(\Delta c>Uc)) \tag{7}$$

When the two alternatives are indifferent, it is acceptable to assume that the choice probability for each alternative will be 0.5. Of course it is possible to give it a certain constant value different from 0.5.

From eq.(5), eq.(6) and eq.(7), the probability that alternative 1 will be chosen is formulated as follows.

$$P(A=1)=\alpha_1 \mathrm{Prob}(\Delta t>Ut)+\alpha_2 \mathrm{Prob}(\Delta c>Uc)(1-\mathrm{Prob}(\Delta t>Ut)) + \frac{1}{2}(1-\mathrm{Prob}(\Delta t>Ut))(1-\mathrm{Prob}(\Delta c>Uc)) \tag{8}$$

The threshold values(Ut,Uc) are assumed to be random variables with log-normal distribution. Then the probability that the difference of time or cost is bigger than the value of threshold is given by eq.(9).

$$\mathrm{Prob}(\Delta x>Ux)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{Yx} e^{-\frac{t^2}{2}}dt \tag{9}$$

where $Yx=a_x \log\Delta x+b_x$

$a_x=\frac{1}{\sigma_x}$

$b_x=-\frac{\mu_x}{\sigma_x}$

μ_x: expected value of log(Ux)

σ_x: standard deviation of log(Ux)

In eq.(8), unknown parameters are a_t, b_t, a_c, b_c. We use the maximum-likelihood method to estimate these parameters.

3. Description of the Data Set

3.1 Source of Data

The data source for this study is the home interview survey of non-grocery shopping trips in 1982, conducted in the suburban residential areas, 20km from the downtown of Tokyo. The destinations of shopping trips were 7 areas including the downtown of Tokyo, the suburban shopping malls and the neighborhood shopping area.(Fig.2) The data obtained in this survey comprise of 1026 reported non-grocery shopping trips by 798 individuals. The outline of the survey is shown in Table-1.

3.2 Profile of Shopping Trips in the Suburban Area of Tokyo

The cross tabulations showing the characteristic of each shopping area are as in Fig.3(the percentage share of articles that the people purchased in each shopping area), Fig.4(the frequency of shopping trips to each destination), Fig.5(value of purchase in one trip), and Fig.6(modal split of the shopping trips). Looking at these tables, the different characteristics of downtown shopping areas in Ginza, Shinjuku and Shibuya, suburban shopping malls in Futago-tamagawa and Tama-plaza and the community shopping centers can be easily understood.

4. The Shopping Destination-choice Models

4.1 Logit Model

There are many studies in which the Logit model is used as a shopping destination-choice model[7][8][9]. In these studies, it is reported that the Logit model has sufficient goodness-of-fit and is applicable to the prediction of shopping destination-choice behavior.

The specified Logit model for the study area is shown in Table-2. This model contains one generic variable(road distance) which represents the travel impedance, and two alternative-specific variables which represent the differences in attractiveness of each shopping center due to goods purchased. The value of overall percent correct and $\bar{\rho}^2$ show that this model is acceptable.

When shopping destination-choice models are used for practical shopping area renewal plannings, there are two types of compulsory variables, namely policy variables and SE variables. The policy variables, such as the number of retail employ, floor space and so on, would characterize the plan. Generally speaking, the boundaries of shopping areas in Japanese cities are not clear, therefore it is very difficult to measure this type of variables precisely. We have tried to incorporate retail employment variable, floor space variable into the model, but these variables have been proven to be insignificant by statistical tests.

The SE variables would provide information about what kinds of customer come to the shopping center. Various SE variables have been tried, leading to a conclusion that all of these are insignificant statistically.

The specified destination-choice Logit model does not include these variables mentioned above. We will apply two models in the next two sections to examine how people evaluate the attractiveness of shopping centers.

4.2 Fuzzy-integral Model

As described above, subjective attractiveness of shopping centers has an important role when people choose their destinations. Fig.7 shows fundamental attributes ratings and overall ratings for seven shopping locations. Low overall ratings for Ginza and Shinjuku in the downtown of Tokyo are mainly due to their high travel impedance. Futago-tamagawa and Tama-plaza, where large scale shopping centers have been developed recently, have high attributes ratings and high overall ratings. In Mizonokuchi area the urban renewal program is unable to be carried on because of oppositions by the local retailers and residents, and the quality of environment of the area is still low. This is the reason why the overall rating for Mizonokuchi is low.

Shoppers are assumed to rate the attractiveness of each shopping location as a function of the fundamental attributes ratings,and to choose one alternative which has the highest overall rating. This decision-making process is formulated using the Fuzzy-integral theory. Estimated parameters in the model are shown in Table-3. The overall percent correct is 81.7%, that is higher than that of the Logit model. The upper boundaries of ratings for travel impedance and atmosphere in shopping locations are higher than those other attributes. This fact is consistent with the above description concerning the characteristics of the seven shopping locations. Thus the Fuzzy-integral model has a high ability to explain shopping destination-choice behavior.

4.3 Lexicographic Model

It seems that in deciding the destinations of non-grocery shopping trips, people consider many factors, i.e., transportation service level, abundance of stores and articles, environment of shopping center and so on. While very complicated structures have to be assumed for destination choice in the Fuzzy-integral model, the Lexicographic model contains very simple assumptions as described in Chapter 2. Many people may possibly have their fixed shopping places and therefore they may be almost unaware of how they actually decide these places. So there is some possibility that simple structure is better than such a model as the Fuzzy-integral model. It has become clear, however, with the data obtained in the survey that the order of properties differs between trip makers(Table-4), and hence the Lexicographic model is considered not applicable in this case.

5. The Mode-choice Models

5.1 Logit Model

A lot of studies on mode choice models show that the Logit model has good fitness and is highly applicable. Table-5 shows the estimated MNL model that has the choice set of four modes, train, bus, car and walk. The attributes of model 1 include only the level of service variables(LOS), while model 2 contains socio-economic characteristics(SE) in addition to LOS. Various SE dummy-variables contribute to improve the model as shown in Table-5. The likelihood ratio and the percent correct of these models are high. The value-of-time that is about ¥28 per minute (about $6 per hour) is reasonable and consistent with other studies in Japan.

Table-6 shows binary Logit models-train and car. Percent correct for car users is similar to the multinomial case. Male trip makers who are 30-50 years old and have car licences and automobiles, evaluate the rail utility extremely low when they go shopping with other companies at weekend.

5.2 Comparison of Three Models

As the three models have different ways of introducing SE variables into the models, as mentioned later, we go about by first using only two generic-variables(travel time, cost) as shown in Table-7 in order to make the models comparable. Although the percent correct index is not the best one for the Logit model, it is the only common index available for comparing these three models. The percent correct of the three models are approximately equal. Strictly speaking, however, the Lexicographic is best and the Logit is worse than the Fuzzy. It would seem then that the Logit has the lowest goodness-of-fit. Nevertheless it is so easy to introduce SE variables into the Logit model to improve this goodness-of-fit. Table-5 and 6 show the fact that SE variables introduced into the Logit model raises the goodness-of-fit of the model remarkably.

5.3 Introduction of SE Variables into Each Model

There are different ways in introducing SE variables into each model. As for the Logit model there are three ways as follows;

1) Introducing them into the model as alternative-specific variables.
2) Introducing them into the model as combination variables of LOS and SE attributes.
3) Segmenting the population with homogeneous value of SE variables.

In the first case, LOS and SE variables constitute an utility function with independent attributes,while in the second case there is another idea to assume that the parameters of LOS variables are affected by SE variables. In the third case, the model is calibrated using data from a segmented group of samples. The purpose of this method is to get parameters with lower standard errors, but the model estimated with segmented data are not always better than the models estimated using pooling data.

For the Fuzzy-integral model the following two ways are possible;
1) Estimating different membership functions for each group that has different values of SE variables.
2) Segmenting the population with homogeneous SE variables.
Membership functions estimated in our analysis for various groups which have different SE characteristics are very similar to each otehr. It has become clear that method 1) is not practical. The segmentation method has the same difficulties as in the Logit model.

For the Lexicographic model there are no other ways than to segment the population.

From the above discussions, it is clear that the Logit model is superior to the other two models as far as the introduction of SE variables is concerned. The introduction of SE variables into the mode-choice Logit models remarkably improves the goodness-of-fit of the models. This fact has been stated in section 5.1.

The Fuzzy-integral model and the Lexicographic model show similar goodness-of-fit to that of Logit model, when they are applied to mode-choice behavior with two LOS variables. It is very difficult, however, to introduce SE variables into the former two models. It is necessary to overcome this weakpoint of these models.

6. Conclusion

The ability of disaggregate travel demand models for predicting shoppers' behavior renders them very useful in the decision-making process in such controversial projects as the redevelopments of shopping centers in Japan.

We have conducted a small size home interview survey in a suburban residential area of Tokyo to analyse the non-grocery shopping trip behavior in Japan. Using this data, three kinds of model-Logit, Fuzzy-integral and Lexicographic model-have been calibrated for destination-choice and mode-choice, and the usefulness of each model has been discussed.

Regarding the mode-choice model for shopping trips, it can be said that the Logit model has high applicability in Japanese cities.

Because of complicated landuse in Japan, we have some difficulties in modelling destination-choice. The Fuzzy-integral model appears to have good fitness as a destination-choice model, but further improvements of this model are necessary with regard to the estimation of the membership functions, the introduction of SE variables and the effective estimation process of parameters.

References

[1] L.A.Zadeh,"Fuzzy Sets", Information and Control 8, 1965 pp.338-353

[2] M.Sugeno,"Teory of Fuzzy Integrals and its Aplications," Doctorial thesis, Tokyo Institute of Technology, 1974

[3] H.Ishida, "Analysis of Human Beings' Recognition, Evaluation and Decision-Making Process for Transportation Planning,"(in Japanese), Doctorial thesis, Tokyo University, 1982

[4] S.Morichi et al.,"Study on mode-choice models in Metropolitan area,"(in Japanese), Transport Economics Research Center, 1978

[5] P.C.Fishburn,"The Foundations of Expected Utility," Reidel, 1982

[6] H.C.W.L.Williams and J.D.Ortuzar,"Behavioral Theories of Dispersion and the Mis-Specification of Travel Demand Models," Transp. Res-B, Vol.16B, No.3, 1982, pp.167-219

[7] F.S.Koppelman,"Destination Choice Behavior for Non-Grocery-Shopping Trips," TRR 673, 1978, pp.157-165

[8] M.E.Richards and M.E.Ben-Akiva,"A Disaggregate Travel Demand Model," Saxon House, 1975

[9] T.A.Domencich and D.McFadden,"Urban Travel Demand: A Behavioral Approach," North-Holland, 1975

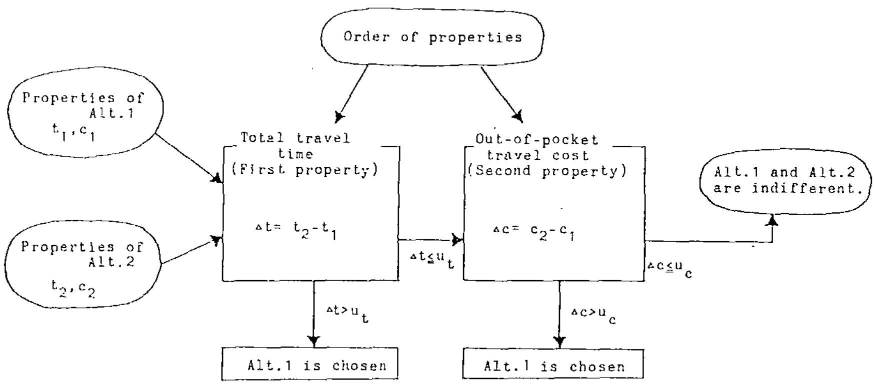


Fig.-1 Decision-making process assumed for the Lexicographic model

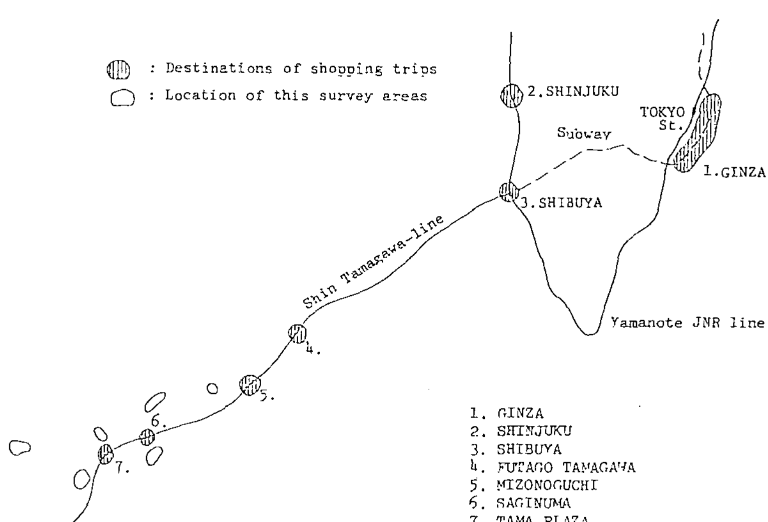


Fig.-2 Survey areas and destinations of shopping trips

Table-1 The outline of the survey

Date	11th-18th Dec. 1982
Survey areas	the suburban of Tokyo
Method of survey	home interview survey random sampling
Sample size	1026 trips
Major survey items	1.Socio-Economic characteristics · sex,age · number of persons in the household · house income · residential location · work location · driver's license 2.shopping trip behavior · shopping frequency · mode used for shopping · parkig spaces availability at shopping locations · value of purchase

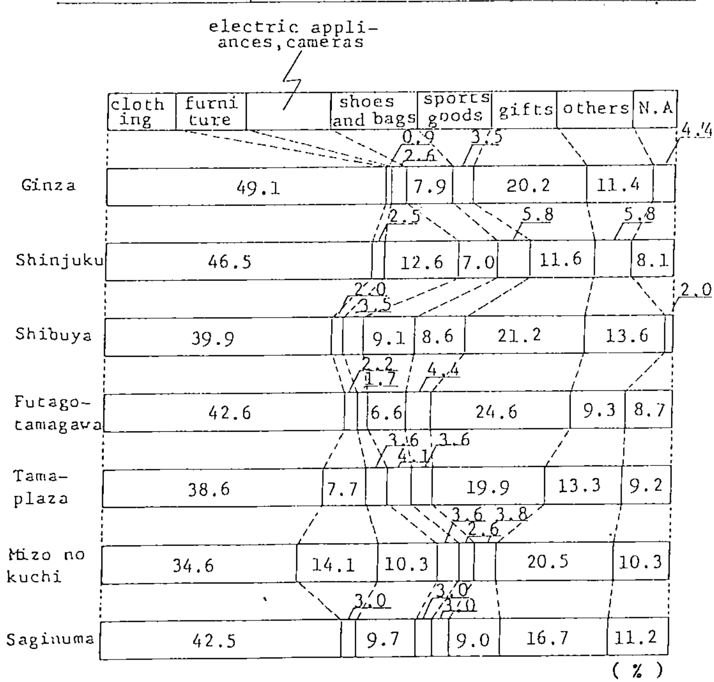


Fig.-3 Articles purchased in each shopping areas

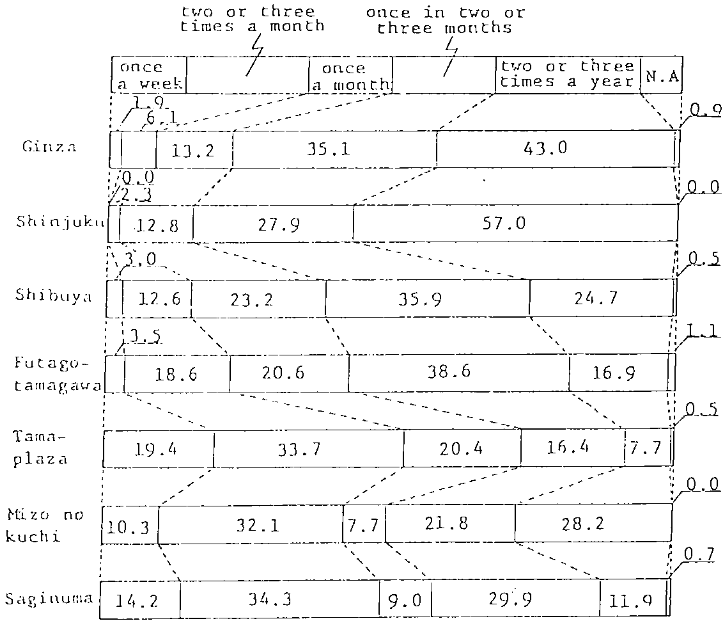


Fig.-4 Frequency of shopping trips

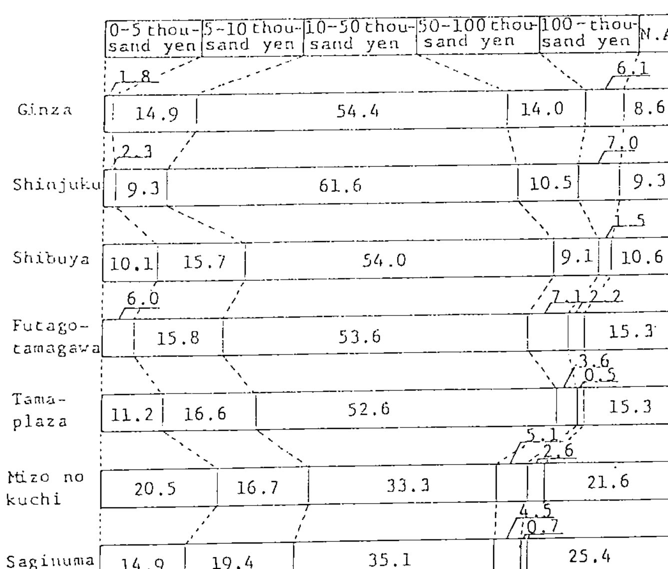
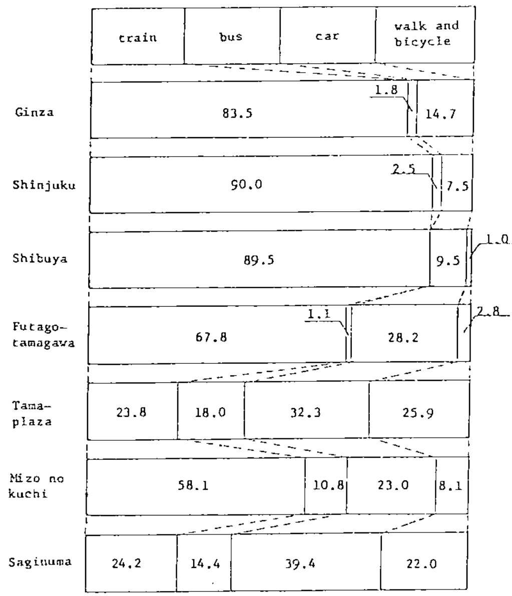


Fig.-5 Value of purchase in one trip



Fog.-6 Modal split of shopping trips

Table-2 Estimation results of MNL,model(destination-choice)

Independent Variable	Estimated Coefficient	t-statistic
Road distance	-0.02303	5.72
Article dummy[1] for GINGA	0.1854	0.52
Article dummy for SHINJUKU	-0.07218	0.17
Article dummy for SHIBUYA	-0.2066	0.73
Article dummy for FUTAGO TAMAGAWA	-0.1934	0.65
Article dummy for MIZONOGUCHI	-0.1064	0.27
Article dummy for SAGINUMA	0.7392	2.11
Shopping center dummy[2] for GINZA	1.252	2.72
Shopping center dummy for SHINJUKU	0.3261	0.72
Shopping center dummy for SHIBUYA	0.4860	1.58
Shopping center dummy for FUTAGO TAMAGAWA	1.652	4.01
Shopping center dummy for MIZONOGUCHI	-3.674	4.83
Shopping center dummy for SAGINUMA	-12.61	0.51
GINZA specific constant	2.707	2.94
SHINJUKU specific constant	2.603	3.11
SHIBUYA specific constant	2.714	4.42
FUTAGO TAMAGAWA specific constant	-0.3008	0.68
MIZONOGUCHI specific constant	0.7410	2.47
SAGINUMA specific constant	0.1168	0.45
χ^2	313.0	
$\bar{\rho}^2$	0.151	
Percent correct	35.5	
Sample size	547	

[1]Variable is 1 for clothing and gifts and 0 otherwise
[2]Variable is 1 for department store and 0 otherwise

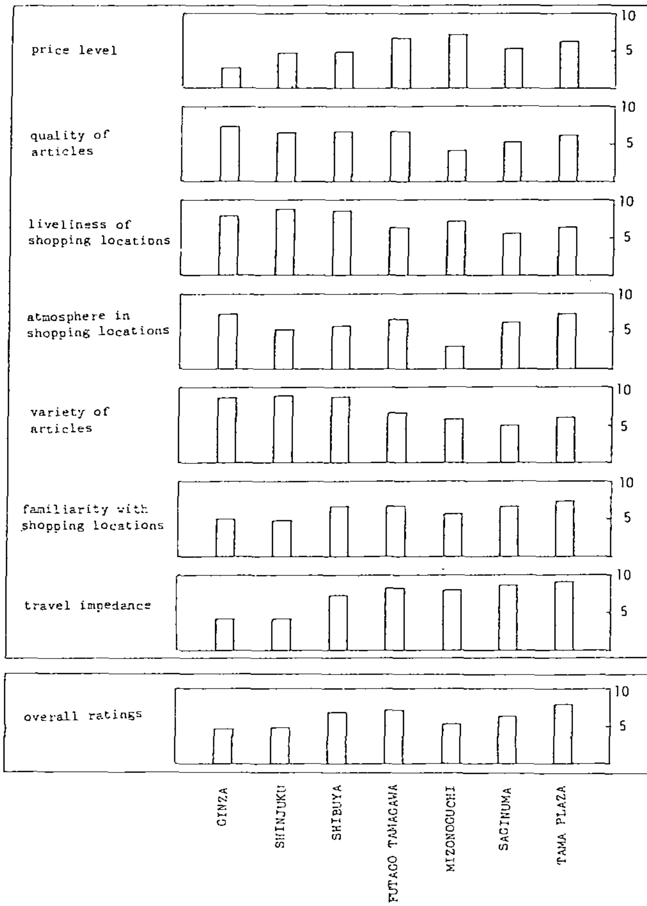


Fig.-7 Average ratings of each shopping center

Table-3 Estimation results of shopping deatination-choice Fuzzy-integral model

λ	Upper boundary of rating $G_\lambda(x_j)$ $j=1,\cdots,7$							overall percent correct
	price level	quality of articles	liveliness of shopping locations	atmosphere in shopping locations	variety of articles	familiarity with shopping locations	travel impedance	
-1.0000	0.00011	0.13628	0.00680	0.97597	0.00302	0.21848	0.99999	81.7 (%)

Table-4 Variation of properties ordering within samples

The ranking of importance for three properties : 1,2 and 3*	Ratio
1 > 2 > 3	32.5%
1 > 3 > 2	25.6%
2 > 1 > 3	16.3%
2 > 3 > 1	11.6%
3 > 1 > 2	7.0%
3 > 2 > 1	7.0%
	100.0%

*Properties : 1 = price level
2 = liveliness of shopping locations
3 = atmosphere in shopping locations

Table-5 Estimation results of MNL.model(mode-choice)

Independent Variable	MODEL 1 Estimated Coefficient	MODEL 1 t-statistic	MODEL 2 Estimated Coefficient	MODEL 2 t-statistic
Total travel time(min)	-0.1286	6.36	-0.1732	6.41
Out of pocket travel cost(yen)	-0.005742	3.99	-0.006186	3.63
A Day of the week dummy[1] for train			-0.1431	0.19
A Day of the week dummy for car			-0.6865	0.98
A Day of the week dummy for bus			1.566	1.39
Accompany dummy[2] for train			-0.5632	0.75
Accompany dummy for car			-1.715	2.37
Accompany dummy for bus			-1.002	0.95
Age dummy[3] for train			-0.06839	0.08
Age dummy for car			1.104	1.37
Age dummy for bus			-0.7281	0.60
Sex dummy[4] for train			-2.791	2.81
Sex dummy for car			-1.048	1.12
Sex dummy for bus			-1.791	1.44
Employment dummy[5] for train			-0.9493	1.13
Employment dummy for car			-0.4824	0.62
Employment dummy for bus			0.7128	0.65
Car ownership dummy[6] for train			-1.692	1.69
Car ownership dummy for car			3.020	2.79
Car ownership dummy for bus			-3.885	2.90
Train specific constant	-0.1796	0.53	2.448	1.73
Car specific constant	-1.117	4.00	-3.309	2.26
Bus specific constant	0.2653	0.63	3.263	1.79
χ^2	479.1		685.5	
$\bar{\rho}^2$	0.445		0.630	
Percent correct	76.0 (train 90.4,car 23.1, bus 100.0,walk 97.6)		84.0 (train 90.8,car 59.8, bus 97.7,walk 90.5)	
Sample size	495		495	

[1]Variable is 0 for weekend and 1 otherwise
[2]Variable is 0 if person travels with other companies and 1 otherwise
[3]Variable is 1 for 30-50 year-old and 0 otherwise
[4]Variable is 1 for male and 0 otherwise
[5]Variable is 1 if person is unemployed and 0 otherwise
[6]Variable is 1 for carowner and 0 otherwise

Table-6 Estimation results of BL.model(mode-choice)

Independent Variable	MODEL 1 Estimated Coefficient	MODEL 1 t-statistic	MODEL 2 Estimated Coefficient	MODEL 2 t-statistic
Total travel time (min)	-0.09542	4.11	-0.1393	4.47
Out of pocket travel cost (yen)	-0.00588	4.07	-0.006564	3.78
A Day of the week dummy[1] for train			-0.6176	1.84
Accompany dummy[2] for train			-1.136	3.40
Age dummy[3] for train			-0.8233	2.32
Sex dummy[4] for train			-1.185	2.97
Car licence dummy[5] for train			-0.8381	2.45
Car ownership dummy[6] for train			-4.335	5.41
Train specific constant	1.036	6.27	7.111	7.42
χ^2	47.1		186.1	
$\bar{\rho}^2$	0.098		0.400	
Percent correct	75.1 (train 96.6,car 13.7)		84.9 (train 93.1,car 61.8)	
Sample size	393		393	

[1]Variable is 1 for weekend and 0 otherwise
[2]Variable is 1 if person travels with other companies and 0 otherwise
[3]Variable is 1 for 30-50 year-old and 0 otherwise
[4]Variable is 1 for male and 0 otherwise
[5]Variable is 1 for licence owner and 0 otherwise
[6]Variable is 1 for carowner and 0 otherwise

Table-7 Comparision of Logit model,fuzzy-integral model and Lexicographic model.(Binary choice - Train & Car)

	Total travel time	Out-of-pocket travel cost	$\bar{\rho}^2$	% Correct	
Logit	θ_t =-0.1817 (12.42)	θ_c =-0.002733 (3.25)	0.209	74.3	83.2 (Train) / 50.2 (Car)
Fuzzy-integral	$G_\lambda(t)$ =1.0	$G_\lambda(c)$ =0.8381	—	75.1	83.5 (Train) / 50.8 (Car)
	λ = -1.0				
Lexicographic	a_t = 0.4776 (3.23) b_t =-0.7365 (2.65)	a_c = 0.5792 (1.70) b_c =-3.521 (2.18)	0.209	75.5	84.1 (Train) / 51.2 (Car)

( ):t-statistic