MULTIPLE INDIVIDUAL ATTRIBUTE EXPANSION METHOD FOR PERSON TRIP SURVEYS

Hiroshi MIURA
Graduate Student
Doctoral Program in Policy and Planning Sciences
University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
Fax: +81-298-53-5591
hmiura@sk.tsukuba.ac.jp

Haruo ISHIDA
Professor
Institute of Policy and Planning Sciences
University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
Fax: +81-298-53-5591
ishida@sk.tsukuba.ac.jp

Tsutomu SUZUKI
Assistant Professor
Institute of Policy and Planning Sciences
University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
Fax: +81-298-53-5591
tsutomu@sk.tsukuba.ac.jp

Abstract: Sample data are expanded to correspond to the actual population of gender and age in person trip surveys in Japan. However, this method might not adequately reproduce the situation of travel behavior. This study proposes the idea of multiple individual attribute expansion (MIAE) method and applies it to the 1988 Tokyo Metropolitan Person Trip Survey data. The MIAE method tries to reproduce simultaneously the distribution of samples on influencing factors of travel behavior such as gender, age, occupation, driver's license ownership, and car ownership as accurately as possible. In Japan, however, detailed cross-tabulation for individual attribute data cannot be used for sample expansion. Some population data are used as the marginal distribution of the multi attribute distribution of the population. The MIAE method produces a smaller total difference between the expanded number of persons and the actual number compared to the present expansion method. The estimates for the transportation situation are slightly changed.

1. INTRODUCTION

In Japan, person trip surveys are carried out in several metropolitan areas where the population is more than 500,000. The major purposes of person trip surveys in Japan are as follows: to grasp the transportation situation including the relation between the change in socio-economic situation and the change in transportation situation; and to make a master plan for a comprehensive urban transportation system based on the estimated traffic volume.

The first general transportation survey in Japan was the Vehicle Origin-Destination Survey in 1958. The purpose of this survey was to clarify the traffic flow on roads. In 1960, the first Metropolitan Transportation Census was carried out in Tokyo, Nagoya, and Osaka metropolitan areas. The number of users of public transportation such as trains and buses was estimated based on this survey. Thus, surveys on private transportation were independent of public transportation surveys. However, the necessity to get simultaneously the situation of private transportation and public transportation increased. So the system of person trip surveys conducted in the United States was introduced in Japan. In 1967, the first person trip survey was carried out in Hiroshima metropolitan area. In the following year, Tokyo metropolitan area had the first person trip survey. Since then, there have been many person trip surveys enforced in Japan. Figure 1 shows the metropolitan areas where person trip surveys have been carried out until 1988.

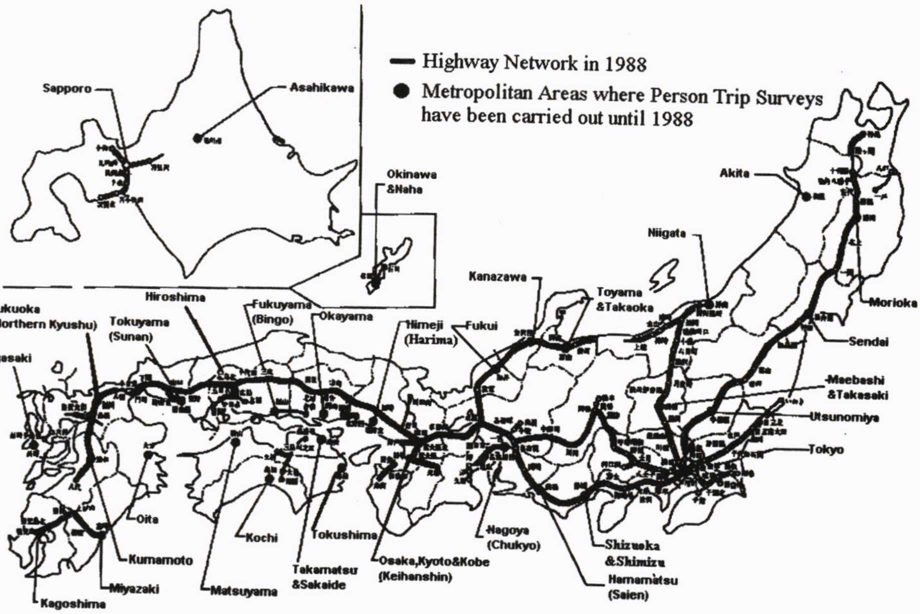

Figure 1. Person Trip Survey Locations

The Basic Resident Registers data, basic population data counted per gender and age class for every city, town, and village have been used in sampling and expanding data for person trip surveys since the first survey. The reasons for it are: the Basic Resident Registers data are good for sampling; and no other data are available for sampling in person trip surveys. At first, sample data were expanded using average sample rate per zone. However, this method could not correspond to the difference in population per individual attribute class. Yamagata (1982) introduced a new expansion methodology with sample rate for each gender and age class. It has been used in Japan until now because this method can reproduce the distribution of population in gender and age accurately.

There are other available population data in Japan. Occupation data are available in Population Census of Japan. The National Police Agency has data on driver's license ownership. The Ministry of Transportation has data on car ownership. Transportation condition will be more effectively estimated using these data than by the present expansion method using the Basic Resident Registers only, although this is less accurate than the present expansion method if we define the expansion as to reproduce exactly the population group from which sample are drawn.

This paper proposes the multiple individual attribute expansion (MIAE) method using these population data and examines its performance and characteristics with the 1988 Tokyo Metropolitan Person Trip Survey (Tokyo PT Survey) data.

The structure of this paper is as follows. Chapter 2 indicates the limitations of the present expansion method. Chapter 3 proposes the idea of the MIAE method. Chapter 4 shows the results of applying this method to the 1988 Tokyo Metropolitan Person Trip Survey data. Finally, Chapter 5 presents the conclusions of this study.

1.1 The 1988 Tokyo PT Survey

The Tokyo metropolitan area (Tokyo, Kanagawa, Saitama, Chiba, and Southern Ibaraki, shown in Figure 2) has a population of 33 million people. It faces many serious problems: urban sprawl, overcrowding, traffic congestion, and so on. To help understand these problems, three person trip surveys were carried out in 1968, 1978, and 1988. In 1998, the fourth person trip survey was performed.

Figure 2. Tokyo Metropolitan Area Map

These person trip surveys consist of two surveys: household survey and screen line survey. In the household survey, sample households are about 3% of all households in the area. The respondents are asked about their individual attributes (gender, age, living place, work place, occupation, industry, driver's license ownership, car ownership, and so on) and about travel behavior in a day (origin place and time, destination place and time, trip purpose, mode choice, parking area, and so on). The present and future transportation conditions are estimated using the household survey data.

2. THE LIMITATIONS OF THE PRESENT EXPANSION METHOD

The present expansion method using the Basic Resident Registers data expands samples to approximate the actual number for each gender and age class. Expansion factors are determined so as to minimize total differences between the expanded samples and the actual population in gender and age. In this method, the expansion factor of certain gender and age class i: E_i is calculated as follows;

$$E_i = N_i / n_i \tag{1}$$

where

N_i : Population number in gender and age class i, and
n_i : Sample number in gender and age class i.

This method can perfectly correspond the expanded samples to the actual population in gender and age class. However, it has been found by travel behavior analysis that individual attributes such as occupation, driver's license ownership, and car ownership have great influence on travel characteristics. It has been also made clear that the number of active travelers such as employees, driver's license holders, and car owners were underestimated by the present expansion method.

For example, Figure 3 shows the differences between the actual numbers of persons and the expanded numbers using the present expansion method for each industry category. It can be seen that the present expansion method underestimates the number of active travelers such as employees in the secondary and tertiary industries. Table 1 represents the trip generation rate for each industry category. Considering these figures, we can say that the present expansion method results in underestimating the number of trips.

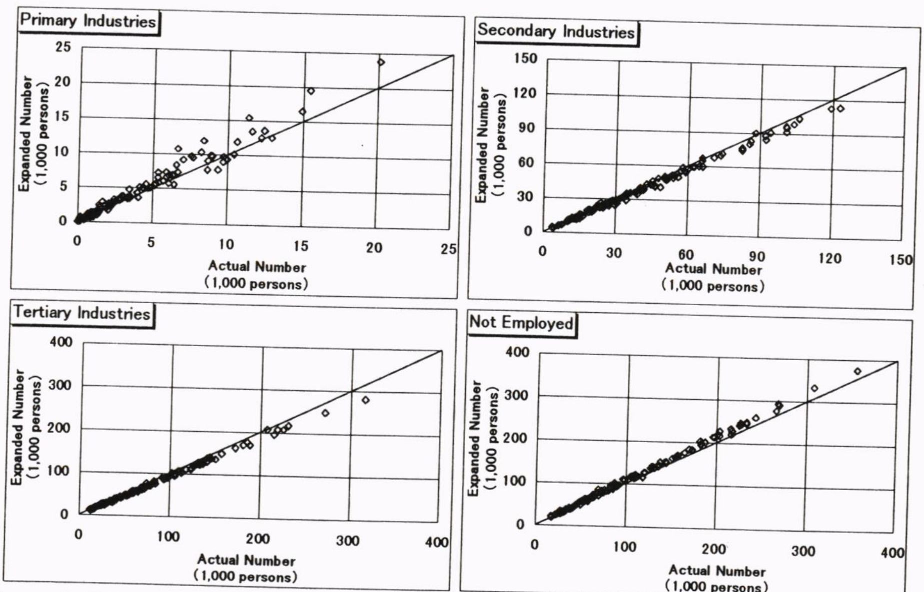

Figure 3. Differences between the Actual Number and the Expanded Number by Industry

Table 1. Trip Generation Rate (trip/person・day) by Industry

Industry	Primary	Secondary	Tertiary	Not Employed
Trip generation Rate	1.956	2.526	2.655	2.223

Moreover, the present expansion method cannot take the differences of the distribution in each individual attribute under consideration. Figure 4 compares the distribution of each individual attribute in all samples with those for males in their 30's. In this figure, the outer circles present the distribution of all samples, and the samples of males in their 30's are shown in inner circles. From this graph we can see that the distribution of each individual attribute such as employment, driver's license ownership, and car ownership are different for each gender and age class. The present expansion method expands samples considering only gender and age. As a result, the distribution of expanded samples in each individual attribute is different from the actual distribution of population such as that shown in Figure 3.

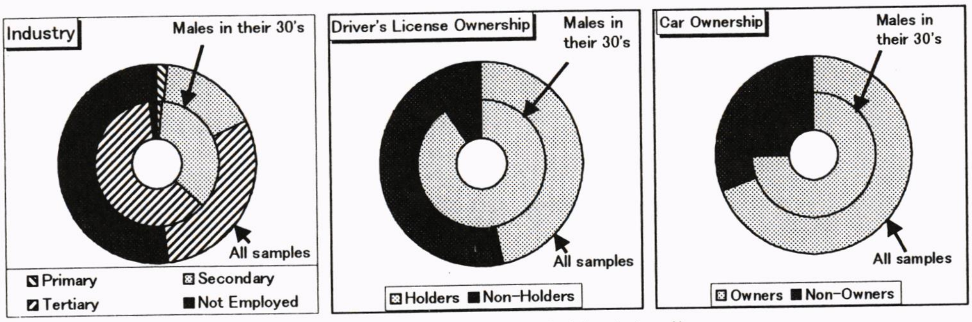

Figure 4. Distribution of Individual Attributes

These results show that the present expansion method might not exactly reproduce the situation of travel behavior such as trip generation, trip distribution, and mode choice. This fact leads to the necessity to consider a new expansion method using multiple individual attributes. Since detailed cross-tabulation for individual attribute data cannot be used in Japan, the Basic Resident Registers data, the Population Census of Japan data, driver's license ownership data, and car ownership data will be used as the marginal distribution data of population.

3. FRAMEWORK FOR MIAE METHOD

3.1 Basic Concept

MIAE method can make the expanded samples close to the actual population in each individual attribute at the same time, although the expanded samples cannot correspond with the actual number in each gender and age class. Expansion factors are calculated to minimize total differences considering the distribution in gender, age, occupation, driver's license ownership, and car ownership.

Figure 5 shows the flowchart of the MIAE method. This section explains how to expand samples using the data of multiple individual attributes.

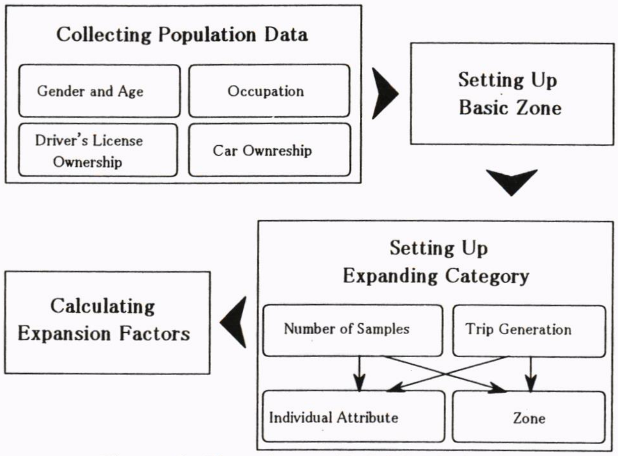

Figure 5. Flowchart of MIAE Method

3.2 Collecting the Population Data in Each Individual Attribute

At first we collect the population data for each city, town, and village for each individual attribute: gender, age, occupation, driver's license ownership, and car ownership. The Basic Resident Registers data are used for gender and age. The Population Census data are used for occupation. The National Police Agency data are used for driver's license ownership. The Ministry of Transportation data are used for car ownership.

The Ministry of Transportation data record the number of cars in each city, town, and village. Since there are no data of population in car ownership in Japan, this data are used as the substitute of car ownership data.

We had the third Tokyo PT Survey in 1988, and two Population Census of Japan were carried out in 1985 and 1990. The estimated population per occupation in 1988 is calculated by using the Population Census data as follows;

$$D_{88} = D_{85} + (D_{90} - D_{85}) \times 3/5 \tag{2}$$

where

D_{88} : Estimated population per occupation in 1988,

D_{85} : Actual population per occupation in 1985 in Population Census, and

D_{90} : Actual population per occupation in 1990 in Population Census.

3.3 Setting Up the Basic Zone for Expansion

The data of distribution in each individual attribute are collected for each city, town, and village. The data in the Tokyo PT Survey are counted for each zone (135 sub zones consisting of some cities, towns, and villages, and 52 main zones consisting of some sub zones). The population data are integrated to correspond to the sub zones of the person trip survey.

Some cities are divided into two or three sub zones in the Tokyo PT Survey. In this case, we correspond the sub zones to the city size.

The expansion factors are calculated for each main zone. However, in some main zones, there are sub zones whose sampling rates are different from each other. In this case, the main zone becomes two zones: a) composing of the sub zones with same sampling rate and b) the other sub zones.

3.4 Setting Up the Expanding Category of Individual Attribute and Zone

At first we set up the expanding category of individual attribute. Sample data are divided into categories for each individual attribute: gender and age class into thirty categories (fifteen male and fifteen female categories with 5-year age groupings), occupation into two categories (employed and unemployed), driver's license ownership into two categories (holders and non-holders), and car ownership into two categories (owners and non-owners).

Some of these categories are integrated considering the number of samples and trip generation rate. If the number of samples in a category is small, a stable expansion factor cannot be calculated. If the categories having different travel characteristics are integrated, the transportation situation cannot be reproduced adequately. It is necessary to integrate categories taking the number of samples and the similarities in the travel characteristics under sufficient consideration. This study integrates categories as follows:

-Considering gender and age class, a category is integrated with another category with similar trip generation rate.

-Considering occupation, driver's license ownership, and car ownership, the categories with small number of samples and similar trip generation rates are integrated.

-After integration, categories will have similar number of samples.

Next, we integrate zones if there are main zones for which the expansion factor cannot be calculated stably. In this study, zones are integrated as follows:

-If there are sub zones in a main zone, the sub zone with the smallest number of samples is integrated with the adjacent sub zone with the highest number of samples.

-For main zones without sub zones, the adjacent main zones with the same sampling rate are integrated.

3.5 Calculating Expansion Factors

The expansion factors are calculated to minimize the sum of squared differences between the expanded samples and the actual population. In MIAE method, the sum of squared differences is calculated for each individual attribute. The formula for the sum of squared differences in this method is shown as follows;

$$F_I = \sum_{i \in I}\left[T_1 \sum_k \left(\sum_{l,m,n} w_{Ij} p_{iklmn} - P_{ik}\right)^2 + T_2 \sum_l \left(\sum_{k,m,n} w_{Ij} p_{iklmn} - P_{il}\right)^2 \right.$$
$$\left. + T_3 \sum_m \left(\sum_{k,l,n} w_{Ij} p_{iklmn} - P_{im}\right)^2 + T_4 \left\{\sum_n \left(\sum_{k,l,m} \frac{w_{Ij} p_{iklmn}}{\sum_n p_{iklmn}} c_{in}\right) - C_i\right\}^2\right] \tag{3}$$

where

F_I : Sum of squared differences between expanded samples and actual population,

I : Main zone number,

i : Sub zone number ($i \in I$),
j : Expansion category ($j = j(k,l,m,n)$),
k : Gender and age category,
l : Occupation category,
m : Driver's license ownership category,
n : Car ownership category,
w_{Ij} : Expansion factor in main zone I and category j,
p_{iklmn} : Sample number of persons in sub zone i and category k,l,m,n,
P_{ik} : Actual number of persons in sub zone i and category k,
P_{il} : Actual number of persons in sub zone i and category l,
P_{im} : Actual number of persons in sub zone i and category m,
c_{in} : Sample number of cars in sub zone i and category n,
C_i : Actual number of cars in sub zone i, and
T_1, T_2, T_3, T_4 : Weight for each individual attribute.

This formula is divided into four parts: gender and age, occupation, driver's license ownership, and car ownership. Expansion category j of a sample corresponds to the combinations of each individual attribute k, l, m, and n. Weights are given for each individual attribute to minimize the differences for each.

The expansion factors are calculated as the solution of the simultaneous equations that are made by differentiating the sum of squared differences (3) with respect to the expansion factors. The simultaneous equations are as follow:

$$\begin{pmatrix} \frac{\partial F_I}{\partial w_{I1}} & \cdots & \frac{\partial F_I}{\partial w_{IJ}} \end{pmatrix}^T = \begin{pmatrix} a_{I11} & \cdots & a_{I1J} \\ \vdots & \ddots & \vdots \\ a_{IJ1} & \cdots & a_{IJJ} \end{pmatrix} \begin{pmatrix} w_{I1} \\ \vdots \\ w_{IJ} \end{pmatrix} - \begin{pmatrix} P_{I1} \\ \vdots \\ P_{IJ} \end{pmatrix} = \mathbf{0} \qquad (4)$$

where

$$a_{Ij_1j_2} = \sum_{i \in I} \left\{ T_1 \sum_k p_{ij_1} p_{ij_2} + T_2 \sum_l p_{ij_1} p_{ij_2} + T_3 \sum_m p_{ij_1} p_{ij_2} + T_4 \sum_n \left(\frac{p_{ij_1} c_{in}}{\sum_n p_{iklmn}} \cdot \frac{p_{ij_2} c_{in}}{\sum_n p_{iklmn}} \right) \right\} \qquad (5)$$

$$P_{Ij_1} = \sum_{i \in I} \left\{ T_1 \sum_k p_{ij_1} P_{ik} + T_2 \sum_l p_{ij_1} P_{il} + T_3 \sum_m p_{ij_1} P_{im} + T_4 \sum_n \left(\frac{p_{ij_1} c_{in}}{\sum_n p_{iklmn}} C_i \right) \right\} \qquad (6)$$

3.6 Expanding Samples

By multiplying the sample data by the expansion factors calculated as the solution of the simultaneous equations (4), we have the population estimates;

$$\hat{p}_{iklmn} = w_{Ij} \cdot p_{iklmn} \qquad (7)$$

where

$\hat{p}_{iklmn}$: Expanded number of samples in sub zone i and category k,l,m,n.

4. APPLYING THE MIAE METHOD TO THE 1988 TOKYO PT SURVEY DATA

This chapter shows the results of applying the MIAE method to the 1988 Tokyo PT Survey data. Expansion factors are calculated for each main zone, and the following are the aggregated results across main zones. To examine the performance and characteristics of this method, we compare three results of expansion:

(1) Present expansion method (Present)
- -Expansion category by gender and age.
- -Expansion factors for only gender and age.

(2) MIAE method A (Method A)
- -Expansion category by gender and age.
- -Expansion factors for all individual attributes (gender, age, occupation, driver's license ownership, and car ownership).

(3) MIAE method B (Method B)
- -Expansion category by all individual attributes (gender, age, occupation, driver's license ownership, and car ownership).
- -Expansion factors for all individual attributes (gender, age, occupation, driver's license ownership, and car ownership).

4.1 Distribution of Expansion Factors

The number of expansion factors for three cases mentioned above are 2921, 1196, and 583 respectively. The distribution of calculated expansion factors for each expansion method, or percentage share is examined and represented in Figure 6. The range of expansion factors by MIAE method is wider than by present expansion method, most of factors are concentrated in between 35 and 45, and that the value of expansion factors is almost stable. This result means that the MIAE method expands the samples properly.

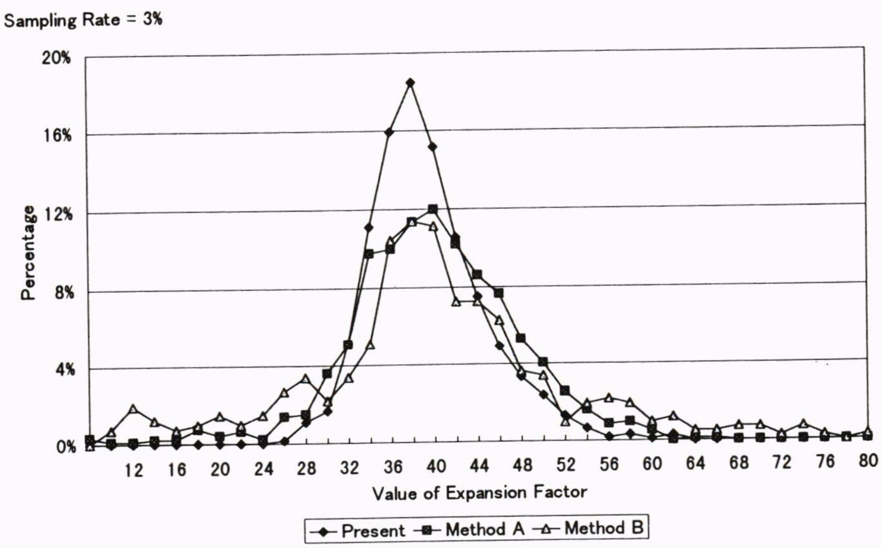

Figure 6. Distribution of Expansion Factors

4.2 Difference between Expanded Number and Population Number

Figure 7 depicts the change in the difference between the population estimates and the actual population by these expansion methods. In this figure, H. Theil's coefficient of inequality: U is used to express the discrepancy:

$$U = \frac{\sqrt{\frac{1}{n}\sum(P_i - A_i)^2}}{\sqrt{\frac{1}{n}\sum P_i^2} + \sqrt{\frac{1}{n}\sum A_i^2}} \tag{8}$$

where

P_i : Estimated number in category i,

A_i : Actual number in category i, and

n : Number of categories for each individual attribute.

If U is close to 0, expanded samples and actual population are very similar. U close to 1 means the expanded samples and actual population are very different.

This figure shows smaller differences between expanded and actual number for MIAE A than the present expansion method for all categories except "number of cars" and "gender and age." MIAE B has smaller differences than the present method for all categories. From this result, we can say that the distribution of estimated population by the MIAE method is more accurate than by the present expansion method. In particular, MIAE B is better than the present method.

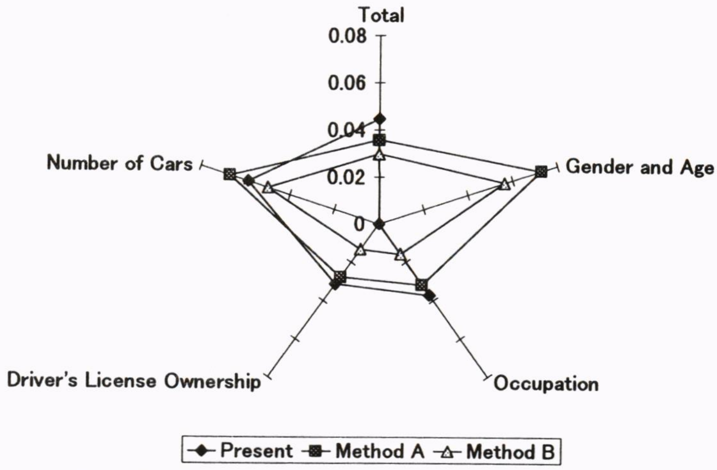

Figure 7. Differences between Expanded Number and Actual Number

4.3 Total Trips and Trip Generation Rate

Table 2 indicates the total number of trips and the trip generation rates of all expanded samples for each expansion method. The total number of trips decreases using MIAE methods. However, MIAE methods bring slightly larger trip generation rates than the present expansion method. The reason for this is the number of expanded samples in MIAE method is less than the present method. Comparing each individual category, the MIAE method is observed to produce larger trip generation rates for people 40 years old or older. Though the number of trips in secondary and tertiary industries increases in the MIAE method, trip generation rates in these industries are very similar.

Table 2. Total Trips and Trip Generation Rate

Expansion Method	Present	MIAE A	MIAE B
Total Trips	74,249,812	73,072,337	74,030,725
Trip generation Rate	2.365	2.390	2.414

4.4 Trip Purpose

Figure 8 represents the percentage share of each trip purpose by expansion method. The purpose "Home – Work" increases by around 500,000 trips and "Home – Private" decreases by nearly 1 million trips using MIAE method. Since the total number of trips in the Tokyo metropolitan area is more than 70 million trips, the composition of trip purpose is only slightly changed by the kind of expansion method.

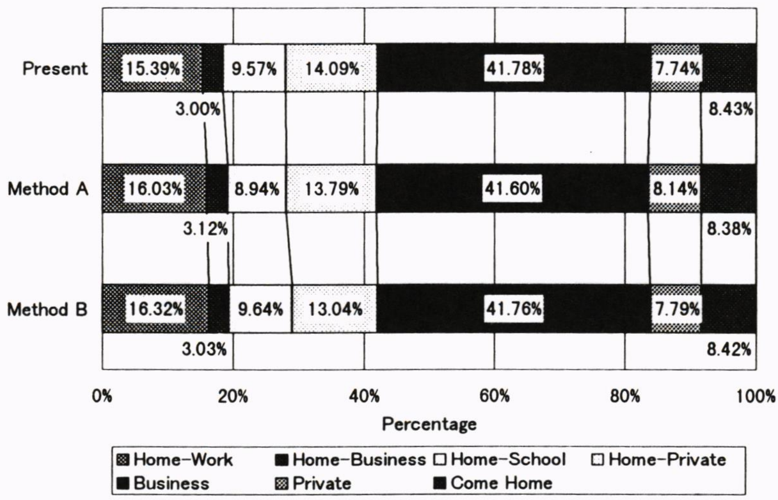

Figure 8. Trip Purpose

4.5 Mode Choice

Figure 9 depicts the percentage share of mode choice for each expansion method. The number of trips using trains and buses increases and that of trips on foot decreases using MIAE method, and the change in the number is less than 500,000. The distribution of mode choice reproduced by each expansion method is hardly changed.

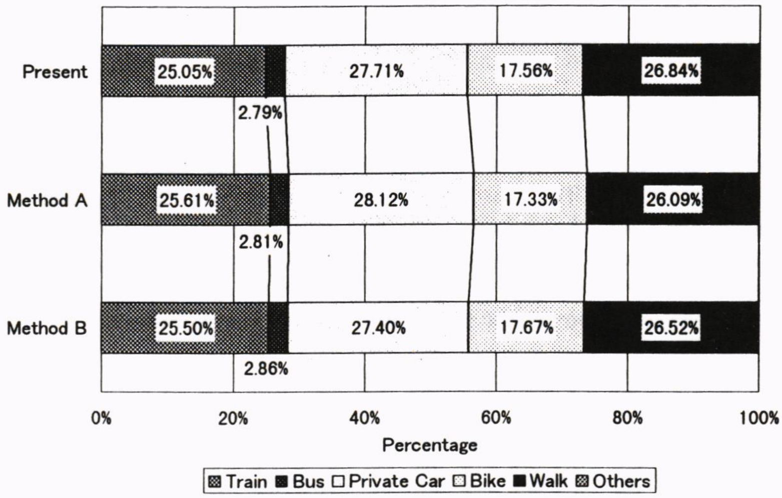

Figure 9. Mode Choice

4.6 Trip Length

Figure 10 indicates the change in distribution of trip length by expansion method. The number of short trips (less than 20 minutes) decreases and that of long trips (more than 20 minutes) increases using MIAE method.

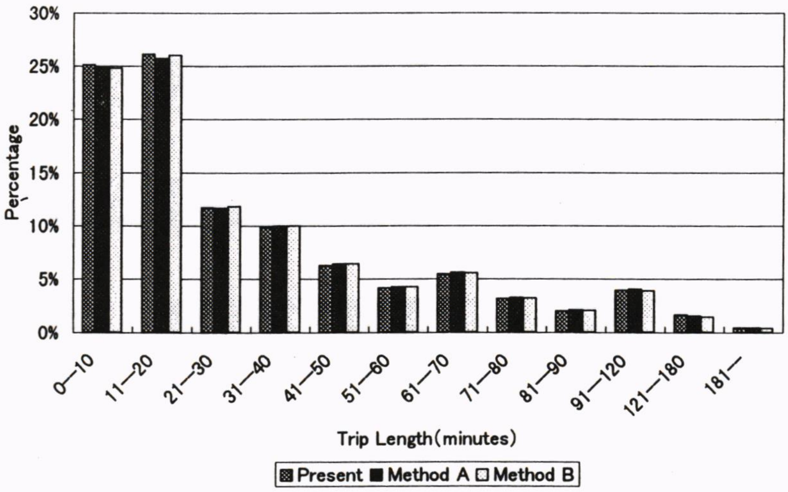

Figure 10. Distribution of Trip Length

Table 3 shows the average trip length for each expansion method. Although the total number of trips decreases, the percentage of longer trips increases as shown in Figure 10. As a result, average trip length is longer using MIAE method compared to the present expansion method.

Table 3. Average Trip Length (minutes)

Expansion Method	Present	MIAE A	MIAE B
Average Trip Length	26.01	26.50	26.19

5. CONCLUSION

This study presents the limitation of the present expansion method which uses only gender and age as key categories. It proposes the idea of MIAE method to expand the sample data considering the more key categories such as gender, age, occupation, driver's license ownership, and car ownership. In Japan, the detailed cross-tabulation for individual attribute data cannot be used for sample expansion. In this method, the Basic Resident Registers data, the Population Census data, the driver's license ownership data, and the car ownership data are used as the marginal distribution data of population.

By applying this method to the 1988 Tokyo PT Survey data, the characteristics of this method are made clear. They are:

-Though the range of expansion factors calculated using MIAE methods is wider than using the present method, most of factors are concentrated around 40.

-The difference between the expanded samples and actual population is less for the MIAE method than the present expansion method.

-The total number of trips is approximately 74 million, which decreases nearly 200,000 trips using MIAE method compare to the present expansion method. However, the trip generation rate slightly increases by MIAE method.

-The characteristics of transportation behavior such as trip purpose, mode choice, and trip length estimated by the present expansion method are slightly changed by the MIAE method.

A difficulty of this methodology is in deciding the expansion categories. If the number of categories is large, expansion factors are not calculated properly. The number of categories in this study is 11; however, expansion factors cannot be calculated stably for some of the main zones. Expansion categories are integrated in such zones. We are now trying a new expansion methodology using Fratar's principle. This is a future task of this study.

ACKNOWLEDGEMENT

The authors would like to thank the Tokyo Metropolitan Transportation Planning Conference for providing us with the data of the 1988 Tokyo Metropolitan Person Trip Survey, and Dr. Hideki Furuya and Mr. Cresencio Montalbo, Jr. for their helpful comments and suggestions.

REFERENCES

Tokyo Metropolitan Transportation Planning Conference (1989) **Report of Tokyo Metropolitan Person Trip Survey in 1988.** Tokyo, Japan. (in Japanese)

U.S. Department of Transportation and U.S. Environmental Protection Agency (1996) Household Travel and Activity Surveys. In U.S. Department of Transportation and U.S.

Environmental Protection Agency, **Travel Survey Manual.** Washington D.C., the United States.

Niitani, Y. (1996) Breakthrough of Transportation Planning by Person-Trip Survey. **Journal of the Japan Society of Civil Engineers, Vol.81-12,** 39-45. (in Japanese)

Yamagata, K. (1982) Differences based on Impossibility of Research on Person Trip Surveys. **Papers of the 4th Research Meeting on Civil Engineering Planning,** 474-481. (in Japanese)

Miura, H., Ishida, H., and Suzuki, T. (1998) A Methodology for Person Trip Data Expansion using Multiple Individual Attributes. **Proceedings of Infrastructure Planning 21(2),** 65-68. (in Japanese)

Theil, H. (Translated by Okamoto, T.) (1964) **Principles of Economics.** Tokyo, Japan. (in Japanese)