Methods of Sampling and Estimation

1. Sampling Method

In this survey, a stratified two-stage sampling method was adopted with the Enumeration Districts of the 2005 Population Census (abbreviated as ED hereinafter) as the first-stage sampling unit, and households as the second-stage sampling unit.

At the first-stage, the whole country was divided into the regions of 47 prefectures, a total of 6,902 sample EDs were selected in those region. The sampling of EDs was done separately for the sample EDs for Questionnaire A and those for Questionnaire B, the number of the sample EDs being 6,513 and 389, respectively. (Some sample EDs in Iwate, Miyagi and Fukushima Prefecture were excluded from the survey because those EDs were devastated by the Great East Japan Earthquake that occurred on March 11, 2011. The number of those EDs in Iwate, Miyagi and Fukushima Prefecture was 25, 35 and 14 respectively. As a result, the number of sample is less than the original.)

At the second stage, 12 households were selected in each of those sampled EDs.

All the members 10 years old and over living in the selected households were enumerated.

(1) Sampling of EDs (First-stage sampling)

Sample EDs for this survey were selected from among all EDs except for

a. Mountain, forest or wilderness districts;
b. Districts with large factories, schools;
c. Districts with public facilities or large hospitals;
d. Districts with prisons or detention houses;
e. Districts of the Self Defense Forces;
f. Districts of foreign military forces;
g. Water surface districts;
h. Isolated islands.

For the sampling, the whole country was divided into the regions of 47 prefectures.

In each prefecture, all EDs were arranged according to the characteristics of EDs, municipality code number and ED number.

Sample EDs were selected by the systematic sampling with unequal probabilities based on population.

(2) Sampling of households (Second-stage sampling)

Ordinarily, EDs have approximately 50 households each on average.

Enumerators visited all households in the sample EDs and drew up the "map of EDs" and the "household lists" before the survey date.

12 sample households were systematically selected for each ED by using a random starting number and a sampling interval designed for each ED.

(3) Reference Day for Time Use

This survey was conducted as of October 20, 2011. However, as for surveying time allocation, two straight days within the nine-day period from October 14 to October 22 were designated for each ED, in order to investigate time allocation by day of the week.

The number of sample by the day of the week was determined allowing for time allocation's dispersion of each day.

2. Estimation Method

The survey results were estimated by the ratio estimation method with a bench-mark of the population by region, sex and age that was estimated from other sources.

The ratio estimates were calculated by the following procedures.

(1) Calculation of the "multiplier for linear estimation"

Firstly, "multiplier for linear estimation of population" is calculated regarding each region (the region corresponds to each prefecture in the calculation on Questionnaire A, and the whole country on Questionnaire B). In other word, "multiplier for linear estimation of population" is the inverse of the sampling ratio.

$$L_{hi} = \frac{1}{m_h} \cdot \frac{Q_h}{Q_{hi}} \cdot \frac{N_{hi}}{n_{hi}}$$

h : Region
i : Index of ED $(i = 1,2,3,\cdots m_h)$
L_{hi} : Multiplier for linear estimation of the i-th ED in the h-th region
m_h : Number of sample EDs in the h-th region
Q_h : Census Population in the h-th region

Q_{hi} : Census Population of the *i*-th ED in the *h*-th region

N_{hi} : Number of household of the *i*-th ED in the *h*-th region

n_{hi} : Number of enumerated household of the *i*-th ED in the *h*-th region

(2) Calculation of linear estimates

Linear estimates of the population by sex and age are calculated regarding each region.

The linear estimates of the population of an ED by sex and age are given by multiplying the enumerated population of the ED by sex and age by the "multiplier for linear estimation" (cf.(1)) of the ED. Then, the linear estimates of the population by region, sex and age are calculated by summing up the linear estimates of each ED within the region.

$$\hat{P}_h = \sum_i^{m_h} L_{hi} \cdot P_{hi}$$

$\hat{P}_h$: Linear estimates of the population in the *h*-th region

P_{hi} : Number of enumerated population of the *i*-th ED in the *h*-th region

(3) Calculation of "multipliers for ratio estimation"

The "multiplier for ratio estimation" is equal to the bench-mark population by region, sex and age* divided by the corresponding linear estimates of the population calculated in (2). Here, the bench-mark population refers to the population as of 1 October 2011 which the Statistics Bureau estimated on the basis of the 2010 Population Census, vital statistics, migration statistics derived from the Basic Resident Registers, etc.

$$R_h = \frac{B_h}{\hat{P}_h}$$

R_h : Mutiplier for ratio estimation in the *h*-th region

B_h : Bench-mark population of the *h*-th region (by region, sex, age)

* Questionnaire A is 15 groups and Questionnaire B is 8 groups.

(4) Calculation of ratio estimates

The procedures of estimating the population with specific characteristics by region, sex and age are as follows.

Firstly, the enumerated population with specific characteristics by sex and age is accumulated within each sample ED in the region.

Then, to multiply the enumerated population with specific characteristics by the "multiplier for linear estimation" makes the linear estimates of the population with specific characteristics in the ED. After that, the linear estimates of the population with specific characteristics by region, sex and age are calculated by summing up the linear estimates of each ED calculated above in the region.

And finally, the ratio estimates of the population with specific characteristics by region, sex and age are given by the product of the "multiplier for ratio estimation" (cf. (3)) and the linear estimates by region, sex and age.

Figures in this report were the aggregation of the ratio estimates of the population by region, sex and age.

(5) Estimating "average time spent on activities per day" and "participation rate in the activity"

"Average time spent on activities per day" and "participation rate in the activity" are computed by the population (or participants in each activity) that is estimated by the above methods.

[Average time spent on activities per day]

$$\begin{pmatrix}\text{Average time spent on Z-activity}\\ \text{in Y-day of the week}\\ \text{who have X-character}\end{pmatrix} = \frac{\begin{pmatrix}\text{Ratio estimates of total time spent}\\ \text{on Z-activity in Y-day of the week}\\ \text{who have X-character}\end{pmatrix}}{\begin{pmatrix}\text{Ratio estimates of population}\\ \text{in Y-day of the week}\\ \text{who have X-character}\end{pmatrix}}$$

"Weekly average" is computed as an average of each day of the week.

$$(\text{Weekly average}) = \frac{(\text{"Monday"} + \cdots\cdots + \text{"Sunday"})}{7}$$

[Participation rate in the activity]

$$\begin{pmatrix}\text{Participation rate}\\ \text{in the Z-activity}\\ \text{who have X-character}\end{pmatrix} = \frac{\begin{pmatrix}\text{Ratio estimates of participants}\\ \text{in the Z-activity}\\ \text{who have X-character}\end{pmatrix}}{\begin{pmatrix}\text{Ratio estimates of population}\\ \text{who have X-character}\end{pmatrix}}$$

3. Standard Errors of Estimates

The standard errors of estimates are approximated by means of the following formula, using four separate estimates made from sub-samples:

$$\hat{\sigma} = \sqrt{\frac{1}{4(4-1)} \sum_{k=1}^{4} (\hat{x}_k - \hat{x})^2}$$

- $\hat{\sigma}$ ↕ Standard error
- k ↕ Index of sub-sample
- $\hat{x}_k$ ↕ Estimate from the k-th sub-sample
- $\hat{x}$ ↕ Estimate from the entire sample