

# Appendix 3 Sampling Method, Estimation Method, and Sampling Errors of Estimates

## 1. Sampling method

The sampling method is a stratified two-stage sampling using the EDs (enumeration districts) of the Population Census as the first-stage sampling units and the dwelling units\* as the second-stage sampling units.

\* A dwelling unit is a structure or a part of a structure intended for habitation of a household, such as a detached house, an apartment, and a living quarter in a dormitory.

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

For each region\*, all EDs are classified into strata according to their characteristic and in each stratum, the sample EDs are selected with the probability proportion to the weight (roughly the number of households in the ED divided by 15). The number of the sample EDs is about 2,900 every month.

The following EDs are excluded from the sampling.

- (a) Reformatory institutions
- (b) Areas resided exclusively by the personnel of the Self-Defence Forces
- (c) Areas resided exclusively by the personnel of foreign armed forces
- (d) Water districts

The data for the inmates of reformatory institutions and the personnel of the Self-Defence Forces are reported by the authorities in charge and are added to the tabulation, whereas in the "Detailed Tabulation", those are excluded from the tabulation.

\* Hokkaido, Tohoku, Southern-Kanto, "Northern-Kanto, Koshin", Hokuriku, Tokai, Kinki, Chugoku, Shikoku, Kyusyu

### (2) Sampling of dwelling units (Second-stage sampling)

For each sample ED, about 15 dwelling units are selected from all dwelling units in the ED by systematic sampling with a constant sampling interval.

All households living in the selected dwelling units, which are approximately 40,000 households in total, are surveyed.

### (3) Sample rotation and structure of sub-samples

In order to improve accuracy of the estimation

on both monthly and annual figures and their changes, the following sample rotation system is applied.

- (a) A sample ED remains in a sample for four consecutive months, leaves the sample for the following eight months, and joins the sample again for the same four months in the following year.
- (b) For each ED, two sets of dwelling units are selected. In the first year of enumeration for the sample ED, the households in the sample dwelling units in the first set are surveyed for the first two consecutive months, and then replaced by the households in the dwelling units of the other set. In the second year, the dwelling units of the first set enter the sample again, and are replaced by those of the other set in the same way as in the first year.
- (c) Under this system, one fourth of the sample EDs and half of the sample households are replaced every month. Three fourths of sample EDs are common from month to month and a half from year to year.

In this system, the whole sample is divided into the following eight sub-samples, each of which forms an independent random sample of the universe. These sub-samples are used in calculation of sampling errors.

The eight sub-samples are represented by A1, A2, B1, B2, C1, C2, D1 and D2.

- A  $i$  ..... EDs beginning in January, May and September
- B  $i$  ..... EDs beginning in February, June and October
- C  $i$  ..... EDs beginning in March, July and November
- D  $i$  ..... EDs beginning in April, August and December
- $i=1$  ..... EDs of the first year
- $i=2$  ..... EDs of the second year

Two sets of dwelling units (set of A2 and C2 or set of B2 and D2, corresponding to the second month in second year) are surveyed of the Special Questionnaire.

## 2. Estimation Method

### (1) Outline of the estimation (Basic Tabulation)

The estimation method for the monthly figures of whole Japan is the ratio estimation with a benchmark of the population by age (15 age groups\*), sex and area (2 areas: metropolitan area consisting of seven largest cities\*\*, and the rest of the country) estimated from the Population Estimates based Population Census(refer to (3)).

\* A change in the method of estimation by 15 age groups (15-19, ..., 80-84, 85-) used in the Labour Force Survey was introduced in 2007.

\*\*Ku-area of Tokyo, Yokohama-shi, Nagoya-shi, Kyoto-shi, Osaka-shi, kobe-shi and Kitakyusyu-shi

The basic formula of the estimation is as follows. (ex. the case for employed person)

$$\begin{aligned} & \text{Ratio estimate of employed person} \\ & = \text{Linear estimate of employed person} \\ & \quad \times \frac{\text{Benchmark population}}{\text{Linear estimate of total population}} \end{aligned}$$

(note) A linear estimation is to estimate the universe by multiplying the enumerated population by the reciprocal of the sampling ratio.

The quarterly and annual averages are the arithmetic mean of the monthly figures.

### (2) Procedure of the estimation (Basic Tabulation)

The estimate of the population having characteristics X is obtained by summing the ratio estimates  $\hat{X}$ s for the sets of sex, age and area concerned.

$\hat{X}$  for a set of sex, age and area is obtained as follows:

$$\begin{aligned} \hat{X} &= \frac{L}{i=1} \frac{1}{m_i} \frac{m_i}{j=1} \frac{\omega_i}{\omega_{ij}} \cdot f_{ij} \cdot x_{ij} \frac{P}{\frac{L}{i=1} \frac{1}{m_i} \frac{m_i}{j=1} \frac{\omega_i}{\omega_{ij}} \cdot f_{ij} \cdot P_{ij}} \\ &= \frac{L}{i=1} \frac{m_i}{j=1} x_{ij} \cdot F_i \frac{P}{\frac{L}{i=1} \frac{m_i}{j=1} P_{ij} \cdot F_i} \end{aligned}$$

$i$  (=1, 2, ....., L) : each section by region and stratum

$j$  (=1, 2, .....,  $m_i$ ) : each sample ED in each section

$x_{ij}$  : Enumerated population (for the sex and age concerned) with characteristic X in  $j$ -th sample ED in  $i$ -th section

$\omega_{ij}$  : Weight for  $j$ -th sample ED in  $i$ -th section

$f_{ij}$  : Reciprocal of the sampling ratio of

dwelling units within  $j$ -th sample ED in  $i$ -th section (=  $\omega_{ij}$ )

$\omega_i$  : Total of the weights for all the EDs in  $i$ -th section

$m_i$  : Number of the sample EDs in  $i$ -th section

$F_i$ (=  $\omega_i / m_i$ ) : Multiplier for linear estimation for  $i$ -th section

$P$  : Benchmark population (for the set of sex, age and area concerned)

$P_{ij}$  : Enumerated population (for the sex and age concerned) in  $j$ -th sample ED in  $i$ -th section

$\frac{P}{\prod_{i=1}^L \prod_{j=1}^{m_i} P_{ij} \cdot F_i}$  : Multiplier for ratio estimation (for the set of sex, age and area concerned)

The estimation method for the 10 regions is a parallel to that for the whole Japan.

For each item, adjustment is made so that the figures for 10 regions add up to the figure for the whole Japan.

### (3) Benchmark population and its revision

“Population Estimates” released by the Statistics Bureau are used as the benchmark population for calculating the survey results.

These population estimates are calculated by using the Population Census counts as the base and by adding the other monthly data on population changes after the Census\*.

Using the latest results of the Population Census as the base, the Population Estimates are revised every five years according to the release of new results of the Census conducted quinquennially. Therefore, the benchmark population for calculating the results of the Labour Force Survey revised every five years.

\* A change in the method of computing the Japanese number of emigrants and immigrants using “the number beyond overseas length-of-stay 91 day” since January 2007.

The benchmark population for the estimation was revised to the 2005 Census-based population estimates in January 2007.

Reference Tab.1 Difference in level of the benchmark revisions (Basic Tabulation)  
(10 thousands, point)

	Population of 15 years old or more	Employed person	Unemployed person	Not in labour force	Unemployment rate
1982 (Bmk 1980 Census)	- 4	- 3	0	- 1	0.0
1987 (Bmk 1985 Census)	+ 7	+ 4	0	+ 3	0.0
1992 (Bmk 1990 Census)	- 11	- 7	0	- 4	0.0
1997 (Bmk 1995 Census)	+ 28	+ 17	+ 1	+ 10	0.0
2002 (Bmk 2000 Census)	- 6	- 4	0	- 2	0.0
2007 (Bmk 2005 Census)	+ 6	- 6	- 1	+ 13	0.0

#### (4) Estimation method for “Detailed Tabulation”

The quarterly and annual averages are the arithmetic mean of the monthly figures.

The estimation method for the monthly figures of whole Japan adopt the ratio estimation with the benchmark of population by age (5 age groups), sex and labour force status (Employed person, Unemployed person and Not in labour force) which estimated in Labour Force Survey\*.

The basic formula of the estimation is as follows.

(ex. the case for employed person on column A in the Special Questionnaire)

Estimate of column A in the Special Questionnaire

= Linear estimate of column A in the Special Questionnaire

×  $\frac{\text{Employed person which estimated in Labour Force Survey}^*}{\text{Employed person which estimated in Detailed Tabulation}}$

Multiplier for linear estimation in Detailed Tabulation is same as Labour Force Survey\*.

\* Results from the Basic Questionnaire

### 3. Sampling error

The magnitude of sampling error varies by the size of estimates, the kind of item and the reference period. Standard errors for estimates which are calculated by using the sub-samples are shown in the following tables.

The inmates of reformatory institutions and the personnel of the Self-Defence Forces are excluded from the calculation.

#### (1) Standard error for the whole Japan by size of estimates (Basic Tabulation)

Tab. 1 Standard error for annual average for the whole Japan

Estimates (10 thousand persons)	Standard error (10 thousand persons)	Relative standard error (%)
5000	12.4	0.2
2000	7.8	0.4
1000	5.5	0.5
500	3.8	0.8
200	2.4	1.2
100	1.7	1.7
50	1.2	2.4
20	0.7	3.7
10	0.5	5.2

The formula used in calculating the relative standard error for annual average of estimates is as follows:

$$\sqrt{\frac{1}{8(8-1)} \sum_{i=1}^8 (\bar{\hat{X}}_i - \bar{\hat{X}})^2} / \bar{\hat{X}}$$

$\bar{\hat{X}}_i$  and  $\bar{\hat{X}}$  respectively refer to the annual average of estimates with characteristic X in the  $i$ -th sub-sample ( $i=1, 2, \dots, 8$ ) and whole sample.

Tab. 2 Standard error for monthly results for the whole Japan

Estimates (10 thousand persons)	Standard error (10 thousand persons)	Relative standard error (%)
5000	24.2	0.5
2000	15.7	0.8
1000	11.3	1.1
500	8.1	1.6
200	5.3	2.6
100	3.8	3.8
50	2.7	5.5
20	1.8	8.8
10	1.3	12.7

The formula used in calculating the relative standard error for monthly estimate is as follows:

$$\sqrt{\frac{1}{8(8-1)} \sum_{i=1}^8 (\hat{X}_i - \hat{X})^2} / \hat{X}$$

$\hat{X}_i$  and  $\hat{X}$  respectively refer to the estimates with characteristic X in the  $i$ -th sub-sample ( $i=1, 2, \dots, 8$ ) and whole sample.

## (2) Standard error for the whole Japan (Basic Tabulation)

Tab. 3 Standard error for major items for annual average (2007 annual average)

Item	Estimates	Standard error	Relative standard error (%)
	(10 thousand persons)		
Labour force	6659	10	0.1
Employed person	6403	11	0.2
Self-employed worker	622	6	1.0
Family worker	236	8	3.2
Employee	5514	9	0.2
Unemployed person	257	2	0.8
Not in labour force	4360	10	0.2
Agriculture and forestry	251	5	2.1
Construction	552	8	1.5
Manufacturing	1165	10	0.9
Information and communications	197	4	2.1
Transport	323	4	1.3
Wholesale and retail trade	1113	4	0.4
Finance and Insurance	155	2	1.0
Eating and drinking places, accommodations	342	3	0.8
Medical, health care and welfare	579	3	0.6
Education, learning support	284	3	1.1
Services(not elsewhere classified)	933	8	0.9

Note) 2007 annual average results do not include the inmates of reformatory institutions and the personnel of the Self-Defence Forces.

Reference tab. 2 Unknown and Unclassifiable for major items (Basic Tabulation)

(10 thousand persons)						
	Labour force status	Hours of work	Status in employment	Number of persons engaged in enterprise	Industry	Occupation
2007 annual average	7	36	30	44	77	70

## 4. Seasonal adjustment methods

- (1) The seasonal adjustment procedure for the labour force series is the U.S. Census Method (X12-ARIMA (X-11 Default)) programme with the control limit for extremes set at 9.8 ~ 9.9, and using standard options for others.
- (2) The seasonal adjustment are calculated on each series independently.
- (3) The seasonally adjusted figures released monthly are calculated by using the predicted seasonal factors, which are estimated through seasonal adjustments of the data up to December of the previous year. On the release

of January results, the seasonal adjustments are computed retroactively by adding new data for twelve months of the previous year.\*

The *reference table 1* in Time Series shows the results of seasonal adjustments using the data up to December 2007.

\* Retroactive computing conducted using the most recent 29 years of original data to produce a final seasonally adjusted estimate, figures are revised the historical seasonally adjusted data for the most recent 10 years.