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, living quarter in a dormitory, hotels, hospitals, construction camps.

(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-Defense 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-Defense 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, Kyushu, Okinawa

> ** Stratification is applied by the industrial and status composition of employed persons and other characteristics in each ED, based on the Population Census. The design of the sampling frame is revised every five years through several phases. The 2015 Census-based design was introduced from May 2018, and completed on August 2019.

(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 (11 regions,) 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.

The basic formula of the estimation is as follows.

[Ex. the case for employed person]

Ratio estimate of employed person

= Benchmark population

$\times$ $\dfrac{\text{Linear estimate of employed person}}{\text{Linear estimate of total population}}$

(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 $\tilde{X}$ s for the sets of sex, age and area concerned.

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

$$\tilde{X} = \sum_{l=1}^{L} \frac{1}{m_l} \sum_{i=1}^{m_l} \frac{w_l}{w_{li}} \cdot f_{li} \cdot x_{li} \cdot r_{li} \frac{P}{\sum_{l=1}^{L} \frac{1}{m_l} \sum_{i=1}^{m_l} \frac{w_l}{w_{li}} \cdot f_{li} \cdot P_{li} \cdot r_{li}}$$

$$= \sum_{l=1}^{L} \sum_{i=1}^{m_l} x_{li} r_{li} F_l \frac{P}{\sum_{l=1}^{L} \sum_{i=1}^{m_l} P_{li} \cdot r_{li} \cdot F_l}$$

l (=1, 2, ……, L) : each section by region and stratum

i (=1, 2, ……, m_l) : each sample ED in each section

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

w_{li} : Weight for i -th sample ED in l -th section

f_{li} : Reciprocal of the sampling ratio of dwelling units within i -th sample ED in l-th section (= w_{li})

w_l : Total of the weights for all the EDs in l- th section

m_l : Number of the sample EDs in l-th section

r_{li} : Adjustment ratio for i -th sample ED in l-th section (change of sampling weight due to the division of ED etc. $0 < r_{li} \leqq 2$)

F_l (= w_l / m_l) :Sampling interval for l -th section

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

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

$\dfrac{P}{\sum_{l=1}^{L} \sum_{i=1}^{m_l} P_{li} \cdot r_{li} \cdot F_l}$:Multiplier for ratio estimation (for the set of sex, age and area concerned)

The estimation method for the 11 regions is a parallel to that 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 2015 Census-based population estimates in January 2017 for Basic Tabulation, and in Jan.-Mar. 2017 for Detailed Tabulation.

Reference Tab. 1
Difference in level of the benchmark revisions
(Basic Tabulation)

(10 thousand persons, point)

	Population aged 15 years old and over	Employed person	Unemployed person	Not in labour force	Unemploy-ment rate
1982 (1981 Yearly Average Results) (benchmark: 1980 Census)	-4	-3	0	-1	0.0
1987 (1986 Yearly Average Results) (benchmark: 1985 Census)	+7	+4	0	+3	0.0
1992 (1991 Yearly Average Results) (benchmark: 1990 Census)	-11	-7	0	-4	0.0
1997 (1996 Yearly Average Results) (benchmark: 1995 Census)	+28	+17	+1	+10	0.0
2002 (2001 Yearly Average Results) (benchmark: 2000 Census)	-6	-4	0	-2	0.0
2007 (2006 Yearly Average Results) (benchmark: 2005 Census)	+6	-6	-1	+13	0.0
2012 (Monthly Results -December 2011-) (benchmark: 2010 Census)	+69	+44	+1	+24	0.0
2017 (Monthly Results -September 2015-) (benchmark: 2015 Census)	+35	+27	0	+7	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 (6 groups), sex, labour force status (Employed person, Unemployed person (ILO 2013) and Not in labour force), status in employment and type of employment which estimated in Basic Tabulation.

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 Question-naire

= Linear estimate of column A in the Special Questionnaire

$\times$ $\dfrac{\text{Employed person which estimated in Basic Tabulation}}{\text{Employed person which estimated in Detailed Tabulation}}$

Multiplier for linear estimation in Detailed Tabulation is same as Basic Tabulation.

3. Sampling error for "Basic Tabulation"

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.

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

Tab. 1 Standard error for annual average for the 2019 whole Japan

Estimates (10 thousand persons)	Standard error (10 thousand persons)	Relative standard error (%)
5000	16.1	0.3
2000	9.7	0.5
1000	6.6	0.7
500	4.5	0.9
200	2.7	1.3
100	1.8	1.8
50	1.2	2.5
20	0.7	3.7
10	0.5	5.1

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

$$\left.\sqrt{\frac{1}{8(8-1)}\sum_{i=1}^{8}\left(\overline{\overline{X}}_i - \overline{\overline{\overline{X}}}\right)^2}\middle/\overline{\overline{\overline{X}}}\right.$$

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

Tab. 2 Standard error for monthly results for the 2019 whole Japan

Estimates (10 thousand persons)	Standard error (10 thousand persons)	Relative standard error (%)
5000	28.2	0.6
2000	17.9	0.9
1000	12.7	1.3
500	9.0	1.8
200	5.7	2.9
100	4.1	4.1
50	2.9	5.8
20	1.8	9.2
10	1.3	13.0

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

$$\left.\sqrt{\frac{1}{8(8-1)}\sum_{i=1}^{8}\left(\widetilde{X}_i - \widetilde{X}\right)^2}\middle/\widetilde{X}\right.$$

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

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

Tab. 3 Standard error for major items for annual average (2019 annual average)

Item	Estimates (10 thousand persons)	Standard error	Relative standard error (%)
Labour force	6886	22	0.3
Employed person	6724	22	0.3
Self-employed worker	531	6	1.1
Family worker	144	1	0.5
Employee	6004	20	0.3
Unemployed person	162	2	1.4
Not in labour force	4197	66	1.6
(Employed person by industry)			
Aguriculture and forestry	207	6	3.0
Construction	499	4	0.8
Manufacturing	1063	10	1.0
Information and communications	229	4	1.7
Transport and postal activities	347	5	1.4
Wholesale and retail trade	1059	7	0.6
Finance and insurance	166	6	3.5
Real estate and goods rental and leasing	129	2	1.2
Scientific research, professional and technical services	240	4	1.7
Accommodations, eating and drinking services	420	4	0.9
Living-related and personal services and amusement services	242	3	1.1
Education, learning support	334	4	1.3
Medical, health care and welfare	843	5	0.6
Services, N.E.C.	455	6	1.2

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

(10 thousand persons)

	Labour force status	Weekly hours of work	Status in employment	Number of persons engaged in enterprise	Industry	Occupation
2019 annual average	9	50	45	86	149	135

4. Sampling error for "Detailed Tabulation"

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

Tab. 4 Standard error for annual average for the 2019 whole Japan (Detailed Tabulation)

Estimates (10 thousand persons)	Standard error (10 thousand persons)	Relative standard error (%)
5000	19.6	0.4
2000	11.9	0.6
1000	8.2	0.8
500	5.6	1.1
200	3.4	1.7
100	2.3	2.3
50	1.6	3.2
20	1.0	4.9
10	0.7	6.7

Tab. 5 Standard error for quarter average for the 2019 whole Japan (Detailed Tabulation)

Estimates (10 thousand persons)	Standard error (10 thousand persons)	Relative standard error (%)
5000	39.3	0.8
2000	23.8	1.2
1000	16.3	1.6
500	11.2	2.2
200	6.8	3.4
100	4.6	4.6
50	3.2	6.4
20	1.9	9.6
10	1.3	13.2

5. Seasonal adjustment methods

(1) The seasonal adjustment procedure for main series is X-12-ARIMA program produced by the U.S. Census Bureau. The other series are seasonally adjusted by X-12-ARIMA (X-11 Default). All series are seasonally adjusted with the control limit for extremes set at 9.8 σ ～9.9 σ , and using standard options for others. For further information on X-12-ARIMA, refer to following website.
URL:https://www.stat.go.jp/english/data/roudou/sa/index.html

(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.* Due to 2012 and 2017 revision of benchmark population, the

input data from October 2005 through December 2016 for seasonal adjustment are not initially released data but comparable time-series data in order to remove gap (excluding rate).

Due to the Great East Japan Earthquake, results for whole Japan are supplementary-estimated figures from March through August 2011.

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

6. Estimation (Time Series Model-based Estimation) by using new statistical methods for the results of the Labour Force Survey (LFS) by prefecture

(1) Back ground

As for the results by prefecture, the estimation methods using Time Series Model-based Estimation have been adopted since May 2006, in order to obtain more stable results. Accordingly, the quarterly average results (model-based estimation) have been* released for reference.

* The yearly average results using ratio estimation (trial) were released since 2002 for reference. However, it was abolished in 2007 due to the full improvement of the time series data of model-based estimation.

(2) Release series

The model-based estimation has released the quarterly and yearly average results by prefecture about the following items since 1997: Labour force, Employed person, Unemployed person, Not in labour force, and Unemployment rate.

(3) Estimation methods

As for the estimation methods for the results by prefecture of the LFS, the Time Series Model-based Estimation which is composed of the following five elements has been adopted.

$$Y(t) = X(t)B(t) + T(t) + S(t) + I(t) + e(t)$$

Observation value: $Y(t)$; Regression: $X(t)B(t)$; Trend: $T(t)$; Seasonal variation: $S(t)$; Irregular variation: $I(t)$; Sampling error: $e(t)$

Notice: Observation value is the survey results value with methods for tabulating the results for the whole country etc (ratio estimation).

Each element shows the following variation:

Regression: It shows a relation between movement of each prefecture and trend of the region which each prefecture belongs to.

Trend: It refers to an aggregated variation of the trend variation which shows longer-term variation with growth of economy etc. and the cyclical variation which has almost constant cycles with business cycle etc. and whose cycle exceeds 12 months.

For example, it is like changes that the number of unemployed persons tends to increase or decrease due to the economic recession and recovery.

Seasonal variation: It refers to a seasonal variation over a 12 month period.

For example, it is like changes that the number of employed persons increases from March to April due to a start of new fiscal year, reaches a peak from May to June, and decreases in latter half of the year.

Irregular variation: It refers to a variation which occurs by sudden event or short-term variation of economy, excluding trend variation, cyclical variation and seasonal variation.

That is to say, it is like changes that output decreases due to a natural disaster such as earthquake or influence of temporary phenomenon such as oil shock.

Sampling error: The LFS has a sample design where around half of the sampled dwelling units remain the same for the comparison with the previous month and the same month of the previous year. Consequently, it is possible to presume that the sampling error has an autocorrelation (it indicates that sampling error of this month is large in case of having a large sampling error in the previous month or the same month of the previous year).

The "Sampling error" refers to the variation pattern regarded as a sampling error and range of variation which are estimated from previous or next time-series data by using the assumed time-series model.

The "**Regression**" obtains a similar variation of trend, and it is possible to regard the aggregation of "**Regression**" and "**Trend**" as an aggregated variation of a trend variation and a cyclical variation.

Using the "**Regression**", time-series variation elements are considered to contain space (area) information, and more multilateral information for estimation is available. The estimation by prefecture using this method is obtained by removing estimation of sampling error ("**Sampling error**") from ratio estimation (the same estimation method as the one for the whole country).

Ratio estimation is used for the results of Hokkaido, Tokyo, Kanagawa, Aichi, Osaka and Okinawa which have relatively-large samples.

(4) Notice for use

In the estimation using Time Series Model-based Estimation, compared with the ratio

estimation, more stable results are obtained by removing a sampling error which is estimated based on time-series model as indicated in (3).

However, the sample of LFS is not designed for tabulation by prefecture (excluding Hokkaido and Okinawa), for its size is small. Using the results by prefecture (model-based estimation) requires attention, because securing enough accuracy of the results is difficult compared with the results for the whole country.

The Time Series Model-based Estimation enables to obtain more stable results by figuring in data after estimation point for model calculation, in addition to data before estimation point. Therefore, quarterly and yearly average results of previous year are recalculated by inserting new results and revised when every January-March quarterly average results are released.