Methods of Sampling and Estimation of Results

1 Sampling Method

In this survey, a stratified two-stage sampling method was adopted with the Enumeration District of the 2020 Population Census (abbreviated as ED hereafter) as first-stage sampling unit, and dwelling as second-stage sampling unit.

At the first stage, about 34 thousand sampled EDs were selected throughout the country, and at the second stage, about 540,000 dwellings were selected in those sampled EDs.

All household members aged 15 and over living in the dwellings sampled in this way were enumerated for the survey.

Besides, consideration was made so that specific households would not be sampled in succession.

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

EDs which are the first-stage sampling units were sampled by the following procedure:

(a) Sampled EDs for this survey were selected from among all the EDs in the country, except the following:

i. districts with mountains, forests or fields
ii. districts with prisons or detention houses
iii. districts of the Self-Defense Forces
iv. districts of foreign military forces
v. water surface districts

(b) All the EDs were stratified into the following 6 Groups according to characteristics based on 2020 Population Census results, etc.

Group 1: EDs with a population of 0
Group 2: EDs with 15 households and under
Group 3: EDs with student dormitories or boarding houses
Group 4: EDs with hospitals, sanatorium or social facilities
Group 5: EDs with employee housing
Group 6: EDs other than the above

(c) For each Group, EDs were arranged under the following standards.

i. Prefecture
ii. Municipal code (upon sampling)
iii. Municipal code (Population Census)
iv. ED's number of the 2020 Population Census

(d) Based on the arrangement in (c), the population of 15 years old and over of all the EDs were totaled for each prefecture, and EDs were sampled with varying probabilities from the totaled population of 15 years old and over.

Table 1 shows the number of sampled EDs for each region selected by the above mentioned (a) to (d)

(2) Sampling of dwellings (second-stage sampling)

Dwellings as the second-stage sampling units were sampled by the following procedure:

(a) All the dwellings were confirmed and a name list was drawn up for each sampled ED, which were arranged in the order of "unoccupied" dwellings to "occupied" dwellings.

(b) Based on the arrangement in (a), the sampling pick up number defines 1, the sampling interval was reached by dividing the number of "occupied" dwellings by 15 and rounding down to the nearest whole number, and dwellings were sampled using the sampling method with equal probability in each sampled ED.

(c) If the number of "occupied" dwellings in the sampled was less than 15, the same process of sampling as in (b) was repeated for the arrangement of dwellings after the "occupied" dwelling that was first sampled, until the number of sampled dwellings reached 15.

2 Estimation Method

The estimation method of survey results was based on the ratio estimation method. For example, the number of persons engaged in work is expressed in the following formula.

$$\text{Number of persons engaged in work} = \frac{\text{Number of persons engaged in work by linear estimation}}{\text{Population by linear estimation}} \times \text{Benchmark population}$$

(Note) Linear estimation is a method of estimating an overall population by multiplying a population obtained in a survey by the

inverse of the sampling rate. In this section, the inverse of the sampling rate is expressed as the multiplier for linear estimation.

In practice, however, the above formula was applied as follows. Populations estimated separately by region, sex, age group and type of household were used as the benchmark population.

$$\text{Number of persons engaged in work} = \frac{\text{Number of persons engaged in work by linear estimation}}{\text{Population by linear estimation}} \times \text{Benchmark population}$$

$$\text{Number of persons engaged in work} = \text{Number of persons engaged in work by linear estimation} \times \frac{\text{Benchmark population}}{\text{Population by linear estimation}}$$

(Note) In this section, the figure reached by dividing the benchmark population by the population by linear estimation is expressed as the multiplier for ratio estimation.

(1) Calculation of multiplier for linear estimation

(a) Categories for calculation

The multiple for linear estimation was calculated for each of the following categories.

i. Group

Of the 6 groups for sampling shown in 1-(1)-(b) above, Group 1, Group 2 and Group 6 were integrated into a single group to give 4 categories.

ii. Region

Prefectures

iii. Sampled ED

iv. Occupancy

This shows whether or not dwellings are occupied at the time of sampling.

(b) Calculation method

The multiplier for linear estimation by group, region, sampled ED, and occupancy (R_{hijk}) is expressed by the following formula.

$$R_{hijk} = \frac{1}{m_{hi}} \cdot \frac{P'_{hi}}{P'_{hij}} \cdot \frac{N_{hijk}}{n_{hijk}} \cdot r_{hij}$$

h : Group (h =1, 2, ･･･4)

i : Region used for linear estimation (prefectures: i=1, 2,･･･, 47)

j : Sampled ED (j =1, 2, ･･･, m_{hi})

k : Occupancy (k =1, 2)

R : Multiplier for linear estimation

m : Number of sampled EDs

P' : Population of 15 years old and over in the Population Census

N : Total number of dwellings

n : Number of surveyed dwellings

r : Adjustment ratio for divided or combined ED

(2) Calculation of linear estimates

(a) Categories for calculation

Linear estimates were calculated for each of the following categories used in ratio estimation.

i. Region (133categories)

a. Prefectural capitals (47)

b. Ordinance-designated cities excluding prefectural capitals (5)

i.e., Kawasaki-shi, Sagamihara-shi, Hamamatsu-shi, Sakai-shi, and Kitakyushu-shi

c. Cities with a population of 300,000 or more except prefectural capitals and ordinance-designated cities (34)

d. Regions other than the above in each prefecture (47)

(Note) Figures in parentheses indicate the number for each category.

ii. Sex (2 categories)

iii. Age group (15 categories)

Five age groups from 15-19 to 80-84, and 85 and over

iv. Type of household (2 categories)

Ordinary households and one-person households

(b) Calculation method

First, survey population by sex, age group and type of household (Q_{hjkl}) were calculated for each group, sampled ED and occupancy. These were multiplied by the multiplier for linear estimation by sampled ED and occupancy (R_{hijk}) and added up within regions

(133 categories) and groups. In this way, linear estimates of population by sex, age group, type of household ($\hat{P}_{hi'l}$) were calculated for each region and group.

Next, the linear estimates for each region and groups calculated in this way were added up within the region, and the linear estimates of population by region, sex, age group and type of household ($\hat{P}_{i'l}$) were calculated.

$$\hat{P}_{hi'l}=\sum_{j}^{m_{hi'}}\sum_{k}^{2}R_{hjk}\cdot Q_{hjkl}$$

$$\hat{P}_{i'l}=\sum_{h}^{4}\hat{P}_{hi'l}$$

h : Group (h =1, 2, ･･･, 4)
i' : Region used for ratio estimation (i' =1, 2, ･･･, 133)
j : Sampled ED (j =1, 2, ･･･, $m_{hi'}$)
k : Occupancy (k =1, 2)
l : Sex, age group, type of household (l=1, 2, ･･･, 60)
$\hat{P}$: Linear estimates
m : Number of sampled EDs
R : Multiplier for linear estimation
Q : Survey population

(3) Calculation of the multiplier for ratio estimation

(a) Categories for calculation

The multiplier for ratio estimation was calculated for each region, sex, age group and type of household, in the same way as (2)-(a) above.

(b) Benchmark population

The benchmark population was the population as of Oct. 1st, 2022, estimated using the data shown below.

i. 2020 Population Census results (Ministry of Internal Affairs and Communications, MIC)
ii. Population Estimate as of Oct. 1st, 2022 (MIC)
iii. Population Estimates as of Oct. 1st, 2022 (Each prefecture)
iv. Number of persons dwelling in camps or ships of the Self-Defense Forces (Ministry of Defense)
v. Number of persons serving sentences in prisons or detention houses and inmates of reformatory institutions (Ministry of Justice)

(Note) iv. and v. were used to estimate the population outside the scope of the survey. Agencies in parentheses indicate doing each survey.

(c) Calculation method

For each region, sex, age group and type of household, the multiplier for ratio estimation ($R'_{i'l}$) was calculated by dividing the benchmark population ($P_{i'l}$) by the linear estimates ($\hat{P}_{i'l}$).

$$R'_{i'l}=\frac{P_{i'l}}{\hat{P}_{i'l}}$$

i' : Region used for ratio estimation (i' =1, 2, ･･･, 133)
l : Sex, age group, type of household (l=1, 2, ･･･, 60)
R' : Multiplier for ratio estimation
P : Benchmark population
$\hat{P}$: Linear estimates

(4) Calculation of estimates

In order to estimate the population with specific characteristics, the survey population with the characteristics by sex, age group, and type of household are calculated by occupancy in each sampled ED first. They are multiplied by the multiplier for linear estimation (cf. (1)) by occupancy in each sampled ED and the multiplier for ratio estimation (cf. (3)) by region, sex, age group and type of household. The values calculated as above are combined in the region to calculate the estimates of populations with the target characteristics in the region by sex, age group and type of household.

They are combined by the region, sex, age group, and type of household as necessary and various figures in this report are obtained.

3 Sampling Error of Estimates

The sampling error was calculated by the sub-sample method.

In the calculation of sampling error by the sub-sample method, several pairs of independent and equal-sized samples are usually selected in the 1st phase of sampling. But in this survey, 6 pairs of

sub-samples were set as the approximation at later sampling, and it was calculated by the following formula using estimates calculated independently by the 6 pairs of sub-samples.

$$\hat{\sigma} = \sqrt{\frac{1}{6(6-1)} \sum_{i=1}^{6} (\hat{X}_i - \hat{X})^2}$$

$\hat{\sigma}$: Standard error of estimates

i : Index of sub-sample (i=1, 2, …, 6)

$\hat{X}_i$: Estimates based on the i-th sub-sample

$\hat{X}$: Estimates based on total samples

The ratio of standard error to the estimates (standard error ratios) calculated as above is shown tables 2 to 4.