

標本設計及び結果の推定方法

1 標本抽出方法

標本抽出方法は、第1次抽出単位を令和2年国勢調査調査区（以下「調査区」という。）、第2次抽出単位を住戸とし、それぞれの抽出単位を層化した後に抽出を行う層化2段抽出法である。

第1次抽出では、全国から約3万4千調査区を抽出し、第2次抽出では、約54万住戸を抽出した。

調査標本は、この方法により抽出された住戸に居住する15歳以上の世帯員全員である。

なお、抽出に当たっては特定の世帯が続けて調査の対象にならないように配慮した。

(1) 調査区の抽出（第1次抽出）

第1次抽出単位である調査区の抽出は以下の手順により行った。

ア 標本調査区は、次のものを除く全国の調査区の中から抽出した。

- i. 山岳・森林・原野地帯等のある区域
- ii. 刑務所・拘留所等のある区域
- iii. 自衛隊区域
- iv. 駐留軍区域
- v. 水面調査区

イ 各調査区について、令和2年国勢調査の結果等に基づく特性により、次の6層に分類した。

- 層1：人口が0の調査区
- 層2：世帯数が15以下の調査区
- 層3：学生の寮・寄宿舎のある調査区
- 層4：病院・療養所及び社会施設のある調査区
- 層5：給与住宅のある調査区
- 層6：上記以外の調査区

ウ 層ごとに、調査区を次の基準により配列した。

- i. 都道府県
- ii. 市区町村コード（標本抽出時）
- iii. 市区町村コード（国勢調査時）
- iv. 令和2年国勢調査調査区番号

エ ウの配列を基に、それぞれ都道府県ごとに全調査区の15歳以上人口を累積し、累積した15歳以上人口に対して確率比例系統抽出により調査区を抽出した。

上記ア～エによって抽出された地域ごとの標本調査区数は表1のとおりである。

(2) 住戸の抽出（第2次抽出）

第2次抽出単位である住戸の抽出は以下の手順により行った。

ア 標本調査区ごとに、調査区に含まれる全ての住戸を確認して名簿を作成し、これを「居住者無」、「居住者有」の順に配列した。

イ アの配列を基に、標本調査区ごとに、抽出起番号を1とし、抽出間隔を「居住者有」の住戸数を15で除し小数点以下を切り上げた値として、等確率系統抽出法により住戸を抽出した。

ウ 抽出された「居住者有」の住戸数が15未満の場合は、最初に抽出された「居住者有」の住戸の次の住戸以降の配列について、抽出住戸数が15になるまでイと同様に再度抽出を行った。

2 結果の推定方法

結果の推定方法は、比推定の考え方に基づいている。例えば、有業者数は次の式で表される。

$$\text{有業者数} = \frac{\text{線型推定による有業者数}}{\text{線型推定による人口}} \times \text{基準人口}$$

(注) 線型推定とは、調査で得られた人口に抽出率の逆数を乗じ、全体の人口を推計することである。この項では、抽出率の逆数を線型推定用乗率と表記する。

ただし、実務上は、上記の算式を次のように用いている。基準人口は、別途推計した地域、男女、年齢階級、世帯の種類別人口を用いた。

$$\begin{aligned} \text{有業者数} &= \frac{\text{線型推定による有業者数}}{\text{線型推定による人口}} \times \text{基準人口} \\ &= \text{線型推定による有業者数} \times \frac{\text{基準人口}}{\text{線型推定による人口}} \end{aligned}$$

(注) この項では、基準人口を線型推定による人口で除した値を、比推定用乗率と表記する。

(1) 線型推定用乗率の算出

ア 算出する区分

線型推定用乗率は、次の区分ごとに算出する。

- i. 層
標本抽出に用いた上記1～(1)ーイに示す6層のうち、層1、層2及び層6を1つに統合した4区分とする。
- ii. 地域
都道府県別
- iii. 標本調査区
- iv. 居住者の有無
住戸の抽出時における居住者の有無とする。

イ 算出方法

層、地域、標本調査区、居住者の有無別の線型推定用乗率 (R_{hijk}) は、次の式で表される。

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

- h : 層 ($h=1, 2, \dots, 4$)
- i : 線型推定に用いる地域 (都道府県: $i=1, 2, \dots, 47$)
- j : 標本調査区 ($j=1, 2, \dots, m_{hi}$)
- k : 居住者の有無 ($k=1, 2$)
- R : 線型推定用乗率
- m : 標本調査区数
- P' : 国勢調査時の15歳以上人口
- N : 総住戸数
- n : 調査住戸数
- r : 修正倍率 (調査区の合併・分割のあった場合の補正值)

(2) 線型推定値の算出

ア 算出する区分

線型推定値は、比推定に用いる次の区分ごとに算出する。

- i. 地域 (133区分)
 - a 県庁所在都市 (47)
 - b 県庁所在都市以外の政令指定都市 (5)
…川崎市、相模原市、浜松市、堺市及び北九州市
 - c 県庁所在都市及び政令指定都市以外の人口30万人以上の都市 (34)
 - d 各都道府県における上記以外の地域 (47)

(注) 括弧内はそれぞれの区分数を表す。
- ii. 男女 (2区分)
- iii. 年齢階級 (15区分)
15～19歳から80～84歳までの5歳階級及び85歳以上
- iv. 世帯の種類 (2区分)
非単身世帯及び単身世帯

イ 算出方法

まず、層、標本調査区、居住者の有無ごとに、男女、年齢階級、世帯の種類別調査人口 (Q_{hijkl}) を求める。これに、標本調査区、居住者の有無別の線型推定用乗率 (R_{hijk}) を乗じ、地域 (133区分) 及び層内で合算する。これにより、地域及び層ごとの男女、年齢階級、世帯の種類別人口の線型推定値 ($\hat{P}_{hi'l}$) が算出される。

次に、こうして算出された地域及び層ごとの線型推定値を地域内で合算し、地域、男女、年齢階級、世帯の種類別人口の線型推定値 ($\hat{P}_{i'l}$) を算出する。

$$\hat{P}_{hi'l} = \sum_j \sum_k R_{hijk} \cdot Q_{hijkl}$$

$$\hat{P}_{i'l} = \sum_h \hat{P}_{hi'l}$$

- h : 層 ($h=1, 2, \dots, 4$)
- i' : 比推定に用いる地域 ($i'=1, 2, \dots, 133$)
- j : 標本調査区 ($j=1, 2, \dots, m_{hi'}$)
- k : 居住者の有無 ($k=1, 2$)
- l : 男女、年齢階級、世帯の種類 ($l=1, 2, \dots, 60$)
- \hat{P} : 線型推定値
- m : 標本調査区数
- R : 線型推定用乗率
- Q : 調査人口

(3) 比推定用乗率の算出

ア 算出する区分

比推定用乗率は、上記(2)ーアと同じ地域、男女、年齢階級、世帯の種類ごとに算出する。

イ 基準人口

基準人口は、次に示す数値を用いて推計した令和4年10月1日現在の人口を用いる。

- i. 令和2年国勢調査結果 (総務省)
- ii. 令和4年10月1日現在推計人口 (総務省)
- iii. 令和4年10月1日現在推計人口 (各都道府県)
- iv. 自衛隊の営舎内又は艦船内の居住者数 (防衛省)
- v. 刑務所、拘留所の被収容者のうち刑の確定している者及び少年院の在院者数 (法務省)

(注) iv及びvについては、調査の対象外となる人口の推計に用いた。(括弧内はそれぞれの調査実施機関)

ウ 算出方法

地域、男女、年齢階級、世帯の種類ごとに、基準人口 ($P_{i'l}$) を線型推定値 ($\hat{P}_{i'l}$) で除して比推定用乗率 ($R'_{i'l}$) を算出する。

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

- i' : 比推定に用いる地域 ($i'=1, 2, \dots, 133$)
- l : 男女、年齢階級、世帯の種類 ($l=1, 2, \dots, 60$)
- R' : 比推定用乗率
- P : 基準人口
- \hat{P} : 線型推定値

(4) 推定値の算出

ある属性を持つ人口を推定するには、まず各標本調査区の居住者の有無ごとに、その属性を持つ男女、年齢階級、世帯の種類別の調査人口を求める。これに上記(1)で求めた標本調査区、居住者の有無別の線型推定用乗率を乗じ、さらに、(3)で求めた地域、男女、年齢階級、世帯の種類別の比推定用乗率を乗じる。こうして算出された値を地域内で合算することにより、目的の属性を持つその地域内の男女、年齢階級、世帯の種類別の人口の推定値を算出する。

これを必要に応じて、地域、男女、年齢階級、世帯の種類について合算して、各種の結果数字を得る。

3 推定値の標本誤差

標本誤差の算出は、副標本方式により行った。

副標本方式による標本誤差の計算では、第1段の標本抽出において独立で均等な大きさの数组の標本を抽出しておくのが本来であるが、この調査では、その近似として事後的に6組の副標本を設定し、この6組の副標本ごとに独立に算出した推定値を用いて、次式により算出した。

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

$\hat{\sigma}$: 推定値の標準誤差

i : 副標本の番号 ($i=1, 2, \dots, 6$)

\hat{X}_i : 第*i*副標本に基づく推定値

\hat{X} : 全標本に基づく推定値

このように計算された推定値に対する標準誤差の割合（標準誤差率）は、表2～4のとおりである。