

## 労働力調査における住戸増加に伴う抽出率変更時の乗率調整方法の検証

小田 大輔<sup>†</sup>  
白松 千明<sup>†</sup>

Examination of weight adjustment when sampling rate changes due to the increase in dwelling units in the Labour Force Survey

ODA Daisuke  
SHIRAMATSU Chiaki

労働力調査では、調査区内の世帯数が大幅に増加した場合、調査を円滑に実施するため、住戸の抽出率変更及び調査区の分割を認めている。その場合には、変更前後の住戸の抽出率の比を線型推定用乗率（以下「線型乗率」という。）に乗じて調査客体の代表性を確保しているが、当該調査区を基にした推定値が他の調査区と比べて過大となった場合、当該調査区の調査結果が他のデータから著しく乖離していると当該調査区の特徴を含む推定値が過大となり、全体の推定値が真の値と大きく乖離する可能性が考えられる。

本稿では、これに対する処理方法として、抽出された1つの調査区が代表する人口の推定値の分布を国勢調査の結果から求め、この分布の上位0.1%点及び四分位数を基にした基準により、外れ値とする労働力調査の調査区を設定して、住戸の抽出率を変更した調査区の線型乗率を調整する方法を検証する。また、本調整方法を実際の調査に適用した場合、結果精度が向上する可能性があることを示す。

キーワード：国勢調査、労働力調査、外れ値、乗率

When the number of households in an enumeration district of the Labour Force Survey has increased significantly, it is allowed to change the sampling rate of dwelling units and to divide the enumeration district in order for enumerators to conduct a survey promptly.

In this case, the original weight is multiplied by the ratio of the sampling rate before and after the change to maintain the representativeness of the samples.

However, if the weight becomes excessively large, the samples of the enumeration district can act as an outlier and the figures of the whole of Japan could be far from the true value.

In this paper, we will examine how to adjust weights by identifying an enumeration district as an outlier based on the top 0.1% point, and the quartile based on the distribution of the linear estimate of population for each enumeration district of the Population Census.

In addition, we will show the adjustment method might have potential to progress the accuracy of the results.

Keywords: Population Census, Labour Force Survey, Outlier, Weight

---

<sup>†</sup> 総務省統計局統計調査部国勢統計課労働力人口統計室

はじめに

労働力調査の標本設計は、国勢調査調査区を第1次抽出単位、住戸を第2次抽出単位とする層化2段抽出となっている。第1次抽出においては、地域、産業及び従業上の地位別等の結果精度を確保するため、国勢調査の結果から得られる調査区に関する詳細な情報を利用し、各調査区を層化したうえで各層の人口に比例するように国勢調査調査区を抽出しており、例えば平成22年国勢調査に基づく層化基準は46層となっている。こうして層化された地域及び層別に調査区は抽出されるため、線型乗率（調査区の抽出間隔）は地域及び層別に異なっており、その最大値（南関東の層符号0403の層（以下、「層0403」という。他の層符号の層についても同様。）：6,072,000）は、最小値（沖縄の層0401：48,000）の126.5倍となっている（総務省統計局（2013））。

層化の基準は、5年に1度、国勢調査の結果を基に見直されている。新たな層化基準に基づく調査区へは、国勢調査の3年後の5月から4年後の8月にかけて段階的に入れ替わる<sup>1</sup>。つまり、新たな層化基準に用いた国勢調査の4年後の7月までは、前回の国勢調査を基に層化された調査区が調査されており、その調査区の層化の情報は、労働力調査の調査時点で約9年前のものとなっている。このため、調査区によっては宅地開発などにより、調査区の状況が国勢調査時点から大きく変化している場合がある。

第2次抽出においては、国勢調査の結果を基に、どの調査区も概ね15世帯<sup>2</sup>が調査されるように住戸の抽出率を設定しているが、大規模共同住宅が建設されるなど、国勢調査後に調査区内の世帯数が大幅に増加した場合でも調査が円滑に実施できるよう、住戸の抽出率変更及び調査区の分割を認めている。その場合は、修正倍率（変更前後の住戸の抽出率<sup>3</sup>の比）を乗じて線型乗率を調整することにより、調査時の調査客体の代表性を確保している。

ただし、1つの建物で複数の国勢調査調査区が設定されている共同住宅が、異なる階数、形状へ建て替えられる場合、団地の再開発など共同住宅の区画が変更された場合<sup>4</sup>は、国勢調査時点の調査区の範囲を特定して調査を行うことが難しいため、調査客体の代表性を確保することが困難となる。

また、世帯数の大幅な増加に加えて、その調査区の属する層の特徴とは異なる特徴への変化が起きた場合、層とは異なる特徴が、大きく調整された線型乗率により推定値に反映されるため、その調査区がいわゆる外れ値として働き、母集団の推定値が真の値から大きく乖離する可能性がある。例えば、老人ホームが取り壊されて企業の大規模な寮が建てられた調査区が調査対象となった場合、そうした変化は母集団においてあまり起きていないと考えられるが、母集団の推定値には、ある特定の産業に属する就業者が増加している結果が強く出るとともに老人が減少している結果が出ることとなる。

調査区状況の変化により、推定値が真の値から大きく乖離している可能性の有無については、国勢調査により事後的な検証が可能である他、調査区状況の変化、算出した推定値、過去の推定値及び当該調査区の推定値等を比較して検証することが考えられるが、調査月の翌月に結果を公表する労働力調査においては、実務上、このような方法で検証を十分に行うことが困難な場合も想定される。また、検証方法や対応方法については、恣意的との

1 過去においては、現在の16か月より短い期間で入れ替えが行われていた。

2 換算世帯が15世帯となるようにしている。換算世帯については1.4参照。

3 調査区を分割した場合は、分割比を加味して調整している。

4 現段階でこうした事例はごく限られているが、団地の再開発は、今後増加すると考えられる。

誤解を招かないよう透明性を確保することが重要であるが、こうした個別の調査区の事由については、調査対象を秘匿する必要性から、個別に行った検証について透明性の十分な確保が困難な場合もあると考えられる。なお、こうした特定の調査区による変動を抑えるため、労働力調査では修正倍率の上限を2倍としている。

本稿は、調査区状況の変化により母集団の推定値が真の値から大きく乖離する可能性に対処し、結果の精度を確保するため、調査区状況が国勢調査時点から大きく変化した労働力調査の調査区について、以下の3点を重視した処理方法を検討することを目的とする。

- ① 影響が極めて大きな調査区のみを安定的に検出し、調整する
- ② その基準と対応方法とが、ユーザーが理解しやすいシンプルなものとする
- ③ 恣意的な判断の余地が予め排除されるよう、基準と処理方法は予め明確に用意する

本稿の構成は以下のとおりである。第1章では、調査区状況が国勢調査時点から大きく変化した労働力調査の調査区について、影響が極めて大きな調査区のみを安定的に検出し、調整する方法を検討する。第2章では、平成25年5月から31年4月までの労働力調査結果を基に、第1章で検討した調整方法と、現行の調整方法による調整結果とを比較する。第3章は、まとめとする。

## 1 影響が極めて大きな調査区の検出方法及びその調整方法の検討

通常、標本から得られた調査結果は、母集団の姿を適切に反映していると考えられるが、調査区状況の変化に対応するため線型乗率を調整したもののうち、線型推定値が極めて大きく外れ値と考えられるものについては、全体の推定値を真の値から大きく乖離させる可能性がある。こうした場合の処理方法として、線型乗率を改めて調整することが考えられるが、その場合は外れ値の検出基準や調整方法について恣意的との誤解を招かないものとする必要がある。

石川・遠藤・白鳥(2010)では、調査で発生するこうした外れ値への対応においては、①正確性の維持、②理解し易さの維持、③作成負担の抑制という3点を留意すべきとしている。また、野呂・和田(2015)では、統計の実務における外れ値の検出に関して、頑健な統計量により外れ値検出のレンジを設定することの重要性が述べられている。

こうした手法を援用し、本章では次の手法を検討する。まず、ある調査区が全体の推定値へ与える影響を表す指標を検討する。次に、母集団である国勢調査の全調査区の分布を表す指標を算出し、これを基に外れ値の基準を設定する。次に、実際の調査において住戸の抽出率を変更した調査区及び分割した調査区が外れ値の基準を超える場合、これを基準まで抑える調査区ごとの修正倍率の上限値を算出する。

### 1.1. ある調査区が全体の推定値へ与える影響

労働力調査は、各調査区の調査対象の人員を数えあげ、線型乗率を乗じて足し上げることで全体の値を推定している。その線型乗率は、調査区が各地域の層別に抽出されるため

---

5 調査区状況の変化により質的な変化が起きていないか、その調査区の調査結果が同一層の他のデータから著しく乖離していないかの判断については本稿では行わないこととする。

層ごとに異なり、例えば全国の総人口（ $P$ ）は次式のとおり推定される。

$$\hat{P} = \sum_k \sum_l \sum_i g_{kl} p_{kli}$$

$$\hat{p}_{kli} = g_{kl} p_{kli} \quad \dots \dots (1)$$

$\hat{P}$  : 全国の総人口推定値  
 $g_{kl}$  : 地域 $k$ 、層 $l$ の線型乗率（調査区の抽出間隔）  
 $p_{kli}$  : 調査区 $i$ の調査対象となる住戸の人口  
 $\hat{p}_{kli}$  : 調査区 $i$ の単位推定人口

つまり、抽出された1つの調査区が代表する人口の推定値（ $\hat{p}_{kli}$ ）（以下、「単位推定人口」という。）は式(1)のとおり地域 $k$ 、層 $l$ の線型乗率に調査区 $i$ の調査対象となる住戸の人口を乗じる（ $g_{kl} p_{kli}$ ）ことで求められ、これはある調査区が全体の推定値に与える影響の大きさを表しているとも考えられる。労働力調査では、国勢調査時点の情報を基に、どの調査区も概ね15世帯<sup>2</sup>が調査されるように住戸の抽出率を設定しているが、調査区内の世帯数が増加して調査対象の人員（ $p_{kli}$ ）が増加した場合、単位推定人口（ $\hat{p}_{kli}$ ）も比例して大きくなることわかる。

また、線型乗率（ $g_{kl}$ ）は層 $l$ のウエイト（ $w_{kl} = \sum_i w_{kli}$ ）を層 $l$ の標本調査区数（ $n_{kl}$ ）で除したものである（ $g_{kl} = w_{kl}/n_{kl}$ ）ことから、単位推定人口（ $\hat{p}_{kli}$ ）は調査区 $i$ の調査対象となる住戸の人口（ $p_{kli}$ ）に1次抽出抽出率の逆数（ $w_{kl}/w_{kli}$ ）及び2次抽出抽出率の逆数（ $w_{kli}$ ）を乗じて推定した、層 $l$ の総人口推定値（ $\hat{P}_{kl,kli}$ ）と層 $l$ の標本調査区数（ $n_{kl}$ ）との比であることわかる。

$$\hat{p}_{kli} = g_{kl} p_{kli} = \frac{w_{kl}}{n_{kl}} p_{kli} = \frac{w_{kl}}{n_{kl}} w_{kli} p_{kli} = \frac{\hat{P}_{kl,kli}}{n_{kl}}$$

$\hat{P}_{kl,kli}$  : 調査区 $i$ から推定した層 $l$ の総人口推定値  
 $n_{kl}$  : 層 $l$ の標本調査区数  
 $w_{kl}$  : 層 $l$ のウエイト

労働力調査では各層の総人口（ $P_{kl}$ ）に比例するように標本調査区数（ $n_{kl}$ ）を設定しているため、調査区 $i$ の人口が国勢調査時点と変わらない場合、その比（ $\hat{P}_{kl,kli}/n_{kl}$ ）、つまり単位推定人口（ $\hat{p}_{kli}$ ）は層に関わらず、ある程度の範囲に収まることが期待される。ただし、標本調査区数（ $n_{kl}$ ）が人口比例的に設定されていない層0403<sup>6</sup>などはこの限りではない。

それでは、国勢調査以降の変化を考慮しない場合、単位推定人口（ $\hat{p}_{kli}$ ）はどの程度の範囲に収まるのであろうか。次節において、国勢調査の結果を基に標本設計時点の分布を確認し、その分布を基に影響が極めて大きな調査区を検出する基準を設定する。

## 1.2. 単位推定人口の算出

式(2)のとおり、平成22年国勢調査の調査区<sup>7</sup>について、国勢調査で得られた調査区の総人口（ $p_{\text{国},kli}$ ）を労働力調査において用いる住戸の抽出間隔（ $w_{kli}$ ）で除したものを労働力調査の調査対象となる住戸の人口とし、これに線型乗率（ $g_{kl}$ ）を乗じることで、式(1)で示した単位推定量（ $\hat{p}_{kli}$ ）を算出する。

$$\hat{p}_{\text{国},kli} = g_{kl} \frac{p_{\text{国},kli}}{w_{kli}} \quad \dots \dots (2)$$

$p_{\text{国},kli}$  : 調査区 $i$ の総人口（国勢調査結果）  
 $w_{kli}$  : 調査区 $i$ の住戸の抽出間隔（住戸の抽出率の逆数）

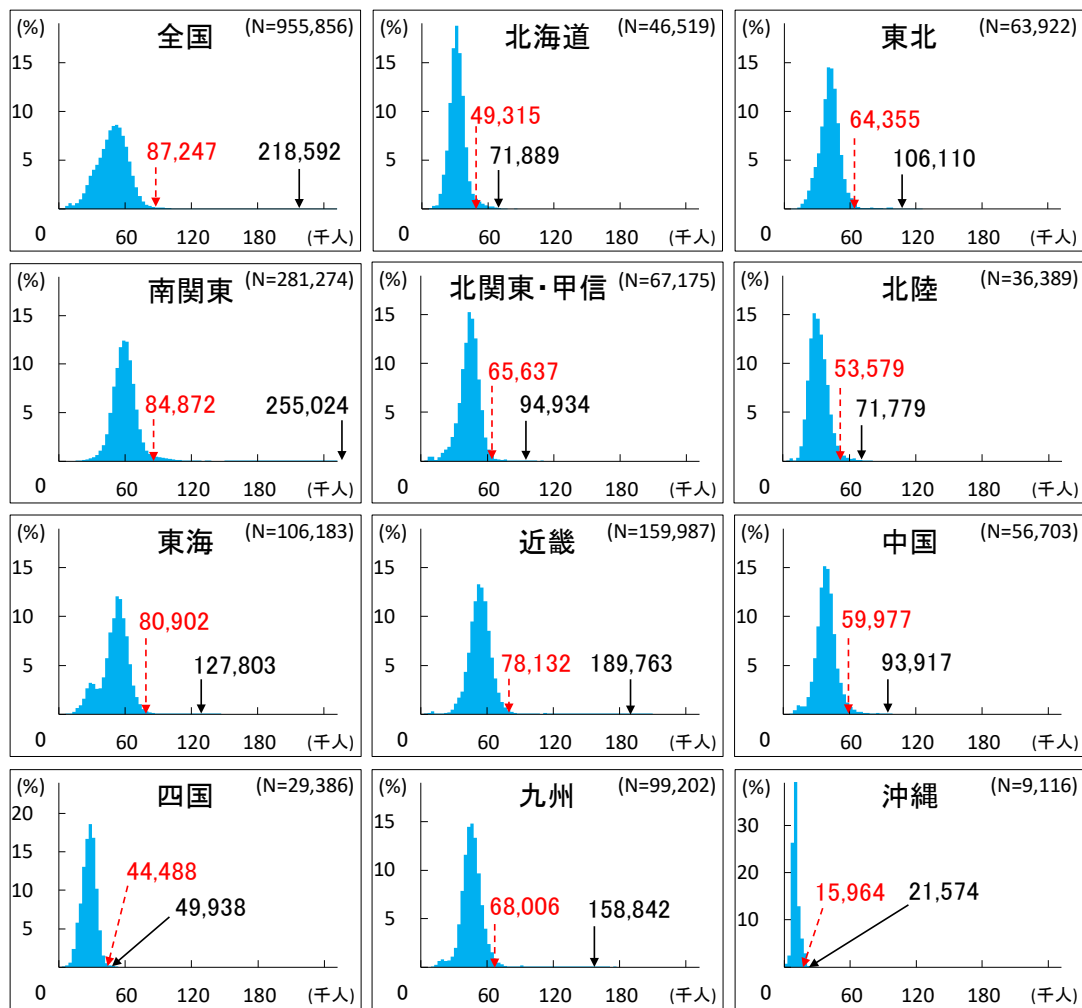
算出した単位推定人口（ $\hat{p}_{\text{国},kli}$ ）の分布を全国及び地域別に確認すると、図1のとおり概

6 老人ホーム、児童保護施設などの社会施設（50人以上の世帯）のある調査区（総務省統計局（2013））。

7 住戸の抽出間隔（ $w_{kli}$ ）が1である層03の調査区は、後述する調査区ごとの修正倍率上限値が極めて大きくなる一方で、世帯の増加により抽出率を変更する調査区の多くは他の層であることから、世帯が増加した場合の修正倍率上限値を求めるという目的に照らし、層02と併せて本集計では除外した。

ね単峰といえる分布であるものの、その分布の範囲は地域ごとに大きく異なることが見て取れる。なお、こうした分布は平成22年国勢調査に特徴的なものではなく、平成27年国勢調査においても同じ傾向がみられた<sup>8</sup>。

また、分布の裾を概観するため、平均+2×標準偏差、箱ひげ図において上側外れ値とされる第3四分位数+1.5×(第3四分位数-第1四分位数)及び上位0.1%点を確認すると、表1のとおり、全国を含めた全ての地域でこれら3つの指標のうち、上位0.1%点が最も大きくなっている。特に全国、南関東、近畿及び九州では、上位0.1%点が平均+2×標準偏差及び箱ひげ図の上側外れ値の2倍以上となっており、こうした調査区の調査結果が同一層の他のデータから著しく乖離している場合、推定結果に極めて大きな影響を与える可能性があると考えられる<sup>9</sup>。



←---: 箱ひげ図の上側外れ値=第3四分位数+1.5×(第3四分位数-第1四分位数) ←: 上位0.1%点

図1. 単位推定人口 ( $\hat{p}_{\text{国},kii}$ ) 別調査区数割合、箱ひげ図の上側外れ値、上位0.1%点

8 参考のため、平成27年国勢調査に基づく単位推定人口 ( $\hat{p}_{\text{国},kii}$ ) 別調査区数割合、地域別上限値 ( $\hat{p}_{k,U}$ )、上位0.1%点を、参考1に掲載する。

9 「推定結果に与える影響が極めて大きい」ことの明確な定義は困難であるが、実務上の経験を踏まえて上位0.1%点を基準とすることとした。

表1. 単位推定人口 ( $\hat{p}_{\text{国},kli}$ ) (人)

地域	平均	標準偏差	第1四分位数	第3四分位数	平均+2×標準偏差	箱ひげ図の上側外れ値(*)	上位0.1%点
全国	49,463	17,588	39,383	58,529	84,638	87,247	218,592
北海道	32,221	7,911	27,450	36,196	48,043	49,315	71,889
東北	41,808	10,427	35,965	47,321	62,661	64,355	106,110
南関東	60,775	17,989	52,801	65,629	96,753	84,872	255,024
北関東・甲信	43,660	9,820	38,598	49,414	63,299	65,637	94,934
北陸	32,369	8,549	26,510	37,338	49,467	53,579	71,779
東海	50,691	13,772	43,907	58,705	78,235	80,902	127,803
近畿	54,512	14,695	47,639	59,836	83,901	78,132	189,763
中国	39,202	9,959	33,621	44,164	59,121	59,977	93,917
四国	27,154	6,446	22,894	31,532	40,047	44,488	49,938
九州	47,200	13,622	40,808	51,687	74,444	68,006	158,842
沖縄	10,514	3,031	8,781	11,654	16,576	15,964	21,574

(\*) 第3四分位数+1.5×(第3四分位数-第1四分位数)

上位0.1%点が箱ひげ図の上側外れ値の2倍以上となっている全国、南関東、近畿及び九州を確認すると、表2のとおり、全国では上位0.37%までが層0403で構成されるなど、全体に占める割合が2%に満たない層0403が上位0.1%点の近辺に分布している。これは、層0403の線型乗率が他の層に比べて大きいためである。

層0403には社会施設を含む調査区が分類されるため、調査対象者の就業状態等は比較的均一であると考えられる。ただし、社会施設の多くは非労働力人口で構成されるものの、ごくまれに就業者で構成される施設も含まれている。このような施設を含む調査区が調査対象となった場合、特に非労働力人口で構成される施設からなる調査区と交代する場合には、いわゆるサンプル要因<sup>10</sup>となり、産業別就業者数等の結果に大きな影響を与える可能性がある。こうした可能性は排除されることが望ましいものの、実務上やむを得ない面もあることから、層0403については、ごくまれに推定結果に極めて大きな影響を与える可能性があるにもかかわらず、大きな線型乗率が設定されていると推測される。

これを踏まえると、こうした特徴を持たない層0403以外の層の調査区の単位推定人口 ( $\hat{p}_{\text{国},kli}$ ) より小さい値を外れ値として検出することは妥当とは考えにくく、例えば平均+2×標準偏差及び箱ひげ図の上側外れ値は、層0403以外の層の調査区と同等の単位推定人口 ( $\hat{p}_{\text{国},kli}$ ) を外れ値として検出してしまうことから、外れ値の基準とすることはできない。

一方で、例えば層0403の調査区において国勢調査後に企業の寮が建てられるなどにより、標本設計時の層化が保たれていない場合、上位0.1%点の単位推定人口 ( $\hat{p}_{\text{国},kli}$ ) の大きさは推定結果に極めて大きな影響を与える可能性があると考えられることから、外れ値の基準とすることは妥当なように思える。ただし、層0403の分布は地域により大きく異なるため、各地域の上位0.1%点を一律に外れ値の基準とした場合、地域によっては外れ値が過度に検出される、またはその逆のケースが発生することが考えられる。

10 層の特徴とは異なる特徴を持つ調査区が選定されること等により、推定値が真の値から外れている(と推定される)こと。

表2. 単位推定人口 (単位別) 別調査区数割合 (%)

単位推定人口(人)	全国		北海道		東北		関東東		北関東・甲信		北陸		東海		近畿		中国		四国		九州		沖縄					
	層0403 の割合	総累計 の割合	層0403 の割合	総累計 の割合	層0403 の割合	総累計 の割合	層0403 の割合	総累計 の割合	層0403 の割合	総累計 の割合	層0403 の割合	総累計 の割合	層0403 の割合	総累計 の割合	層0403 の割合	総累計 の割合	層0403 の割合	総累計 の割合	層0403 の割合	総累計 の割合	層0403 の割合	総累計 の割合	層0403 の割合	総累計 の割合				
0 ~	1.21	98.79	1.00	100.00	1.59	98.41	1.00	100.00	1.36	98.64	1.00	100.00	1.11	98.89	1.00	100.00	1.57	98.43	1.00	100.00	1.57	98.43	1.00	100.00	1.27	98.73	1.00	100.00
6000 ~	0.01	0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7000 ~	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8000 ~	0.01	5.85	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9000 ~	0.03	11.38	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10000 ~	0.02	15.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11000 ~	0.04	17.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12000 ~	0.06	15.74	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13000 ~	0.04	17.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
14000 ~	0.07	11.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
15000 ~	0.05	5.59	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
16000 ~	0.06	5.59	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17000 ~	0.06	2.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18000 ~	0.04	0.72	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
19000 ~	0.05	0.34	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20000 ~	0.07	0.23	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21000 ~	0.05	0.16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
22000 ~	0.03	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
23000 ~	0.05	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
24000 ~	0.02	0.23	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25000 ~	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
26000 ~	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
27000 ~	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
28000 ~	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
29000 ~	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

注 赤枠で囲まれた階級は、後述する地域別上限値を含む階級

### 1.3. 単位推定人口の地域別上限値の算出

前節のとおり、各地域の上位0.1%点を一律に地域別の外れ値の基準とすることは妥当ではないと考えられることから、地域の分布に即した基準を求めるため、いわゆる頑健な統計量と言われている四分位数を用いることを試みる。具体的には、全国の上位0.1%点と第3四分位数との距離と、第1四分位数と第3四分位数との距離の比である定数 $C_U$ を求め、この定数 $C_U$ を各地域の第1四分位数( $\hat{P}_{k,1stQ}$ )及び第3四分位数( $\hat{P}_{k,3rdQ}$ )に当てはめることで、地域別外れ値の基準( $\hat{P}_{k,U}$ ) (以下「地域別上限値」という。)を求めることとする。

$$C_U = \frac{\hat{P}_{0.1\%} - \hat{P}_{3rdQ}}{\hat{P}_{3rdQ} - \hat{P}_{1stQ}}$$

$$\hat{P}_{k,U} = \hat{P}_{k,3rdQ} + C_U(\hat{P}_{k,3rdQ} - \hat{P}_{k,1stQ})$$

$\hat{P}_{0.1\%}$	: 全国の上位0.1%点
$\hat{P}_{1stQ}$	: 全国の第1四分位数
$\hat{P}_{3rdQ}$	: 全国の第3四分位数
$\hat{P}_{k,1stQ}$	: 地域 $k$ の第1四分位数
$\hat{P}_{k,3rdQ}$	: 地域 $k$ の第3四分位数

単位推定人口( $\hat{p}_{\overline{ki}}$ )の全国の分布を基に求めた定数 $C_U$ は8.360となり、これにより求めた地域別上限値( $\hat{P}_{k,U}$ )は表3、図2のとおりとなる。いずれの地域においても、山の分布に応じた一定の距離に地域別上限値( $\hat{P}_{k,U}$ )があることが見て取れ、その値は中央値の3倍から5倍の範囲となっている。分布からの距離とその大きさから、こうした調査区の調査結果が他のデータから著しく乖離している場合、推定結果に極めて大きな影響を与える可能性があると考えられる。以下、これを地域別の外れ値の基準とし、調査区状況が国勢調査時点から大きく変化した調査区について結果を調整する手法を検討することとする。

表3. 地域別上限値 ( $\hat{P}_{k,U}$ ) (人)

地域	地域別上限値 ( $\hat{P}_{k,U}$ )	定数 $C_U$ ※	第1四分位数	第3四分位数	中央値	上位0.1%点
全国	-	8.360	39,383	58,529	49,594	218,592
北海道	109,316	(4.081)	27,450	36,196	31,709	71,889
東北	142,256	(5.177)	35,965	47,321	41,786	106,110
南関東	172,876	(14.764)	52,801	65,629	59,154	255,024
北関東・甲信	139,835	(4.209)	38,598	49,414	44,208	94,934
北陸	127,859	(3.181)	26,510	37,338	31,457	71,779
東海	182,420	(4.669)	43,907	58,705	52,102	127,803
近畿	161,808	(10.652)	47,639	59,836	53,732	189,763
中国	132,300	(4.719)	33,621	44,164	38,792	93,917
四国	103,742	(2.131)	22,894	31,532	27,491	49,938
九州	142,639	(9.849)	40,808	51,687	46,147	158,842
沖縄	35,674	(3.453)	8,781	11,654	9,929	21,574

※ 地域別上限値( $\hat{P}_{k,U}$ )の算出に使用する定数 $C_U$ は、全国のもの(8.360)であるが、参考のため、各地域の第1四分位数、第3四分位数及び上位0.1%点から算出した定数 $C_{k,U}$ を括弧内に掲載した。また、定数 $C_U$ 、 $C_{k,U}$ は無単位数



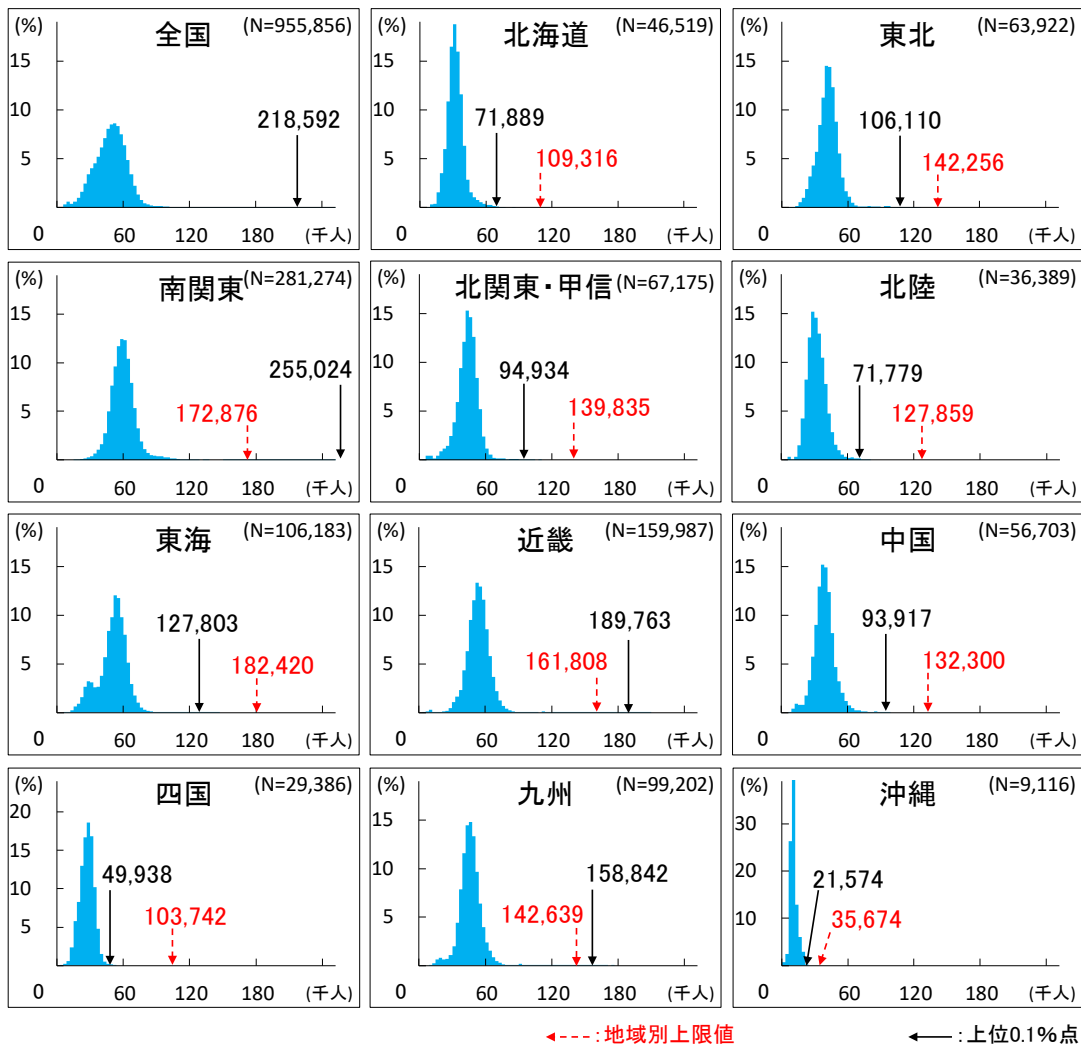


図2. 単位推定人口 ( $\hat{p}_{\text{国},kli}$ ) 別調査区数割合、地域別上限値 ( $\hat{P}_{k,U}$ )、上位0.1%点

#### 1.4. 調査区別の修正倍率上限値の算出方法

労働力調査では、調査区の調査を開始する前月に準備調査を行い、調査区内の全住戸数及びその居住の有無等を確認し、調査対象となる世帯数が調査手順において定めた数を上回る調査区については、調査が円滑に実施できるよう、同じく定められた方法により住戸の抽出率が変更可能となっている。本稿では、統計作成の実務に即した調整方法<sup>11</sup>として、実際の調査結果ではなく、この準備調査の結果を基に修正倍率の調整を行うことを検討する。なお、労働力調査では調査の円滑な実施のため、当初は1調査区あたり15世帯が調査対象となるよう、国勢調査後の世帯の大幅な増加により抽出率を変更する際は1調査区あたり20世帯が調査対象となるよう、それぞれ住戸の抽出率を設定しているが、その際は調査区の世帯数をそのまま用いるのではなく、単身の一般世帯<sup>12</sup>及び施設等の世帯人員<sup>13</sup>を2

11 労働力調査は、調査から公表までの期間が短く、実際の調査結果を基に調整を行ったうえで結果を算出して公表することが困難であると思われる。

12 国勢調査の世帯の種類による分類による。間借り・下宿などの単身者、会社などの独身寮の単身者。

13 寮・寄宿舎の学生・生徒、病院・療養所の入院者、社会施設の入所者、など。

人以上の一般世帯の1/3世帯とする換算世帯数を用いて抽出率を設定している。

調査区別の修正倍率上限値（以下「調査区別上限値」という。）の算出に用いる準備調査時点の単位推定人口（ $\hat{p}_{kli}$ ）は、この換算世帯数を用いて算出することとする。

$$h = h_{2人以上} + \frac{1}{3}(h_{単} + h_{施})$$

$$w_{kli} = \frac{1}{15} h_{kli}$$

$$w'_{kli} = \frac{1}{20} h'_{kli}$$

$h$	: 換算世帯数（小数点以下四捨五入）
$h_{2人以上}$	: 世帯人員が2人以上の一般世帯の世帯数
$h_{単}$	: 世帯人員が1人の一般世帯の世帯数
$h_{施}$	: 施設等の世帯人員
$w_{kli}$	: 調査区 <i>i</i> の住戸の抽出間隔＝抽出率の逆数 （小数点以下切り上げ）
$w'_{kli}$	: 調査区 <i>i</i> の抽出率変更後の住戸の抽出間隔＝抽出率の逆数 （小数点以下切り上げ）
$h_{kli}$	: 調査区 <i>i</i> の国勢調査時点の換算世帯数
$h'_{kli}$	: 調査区 <i>i</i> の準備調査時点の換算世帯数

はじめに、国勢調査結果の調査区*i*の人口（ $p_{\text{国},kli}$ ）に、換算世帯数の比（ $h'_{kli}/h_{kli}$ ）を乗じることで求めた準備調査時点の調査区*i*の総人口を式(2)に当てはめ、住戸の抽出率を変更した調査区の準備調査時点の単位推定人口（ $\hat{p}_{kli}$ ）を式(3)のとおり推定する<sup>14</sup>。なお、線型乗率（ $g_{kl}$ ）には変更前後の住戸の抽出率の比である修正倍率（ $m'_{kli}$ ）を乗じる。

$$\hat{p}_{kli} = g_{kl} m'_{kli} \frac{p_{\text{国},kli} h'_{kli}/h_{kli}}{w'_{kli}} \dots \dots (3)$$

$\hat{p}_{kli}$	: 調査区 <i>i</i> の準備調査時点の単位推定人口
$m'_{kli}$	: 調査区 <i>i</i> の準備調査時点の修正倍率
$p_{\text{国},kli}$	: 調査区 <i>i</i> の総人口（国勢調査結果）

次に、準備調査時点の単位推定人口（ $\hat{p}_{kli}$ ）が地域別上限値（ $\hat{P}_{k,U}$ ）以下となる修正倍率を求める。まず、 $\hat{P}_{k,U} \geq \hat{p}_{kli}$ となる条件は式(4)のとおり表せる。

$$\hat{P}_{k,U} \geq \hat{p}_{kli} = g_{kl} m'_{kli} \frac{p_{\text{国},kli} h'_{kli}/h_{kli}}{w'_{kli}} \dots \dots (4)$$

国勢調査時点で調査区が無人でない場合、式(4)の各変数はいずれもゼロより大きいため、準備調査時点の単位推定人口（ $\hat{p}_{kli}$ ）を地域別上限値（ $\hat{P}_{k,U}$ ）以下とする調査区別上限値（ $m_{kli,U}$ ）は、式(5)のとおり整理できる。

$$m'_{kli} \leq \frac{\hat{P}_{k,U} h_{kli} w'_{kli}}{g_{kl} p_{\text{国},kli} h'_{kli}} = m_{kli,U} \dots \dots (5)$$

以上より、調査区別上限値（ $m_{kli,U}$ ）は、準備調査時点の換算世帯数（ $h'_{kli}$ ）の関数である $\hat{P}_{k,U} h_{kli} w'_{kli} / g_{kl} p_{\text{国},kli} h'_{kli}$ となる。ただし、本式により $1 > m_{kli,U}$ となる調査区別上限値（ $m_{kli,U}$ ）は1とする<sup>15</sup>。

そして、準備調査時点における単位推定人口（ $\hat{p}_{kli}$ ）が、地域別上限値（ $\hat{P}_{k,U}$ ）を超過し

14  $p_{\text{国},kli} h'_{kli}/h_{kli}$ =準備調査時点の調査区*i*の総人口、 $g_{kl} m'_{kli}$ =抽出率変更後の調査区*i*の線型乗率。

15 標本設計時の単位推定人口（ $\hat{p}_{\text{国},kli}$ ）が地域別上限値（ $\hat{P}_{k,U}$ ）を超過している場合、調査区別上限値（ $m'_{kli}$ ）は1未満となる。

た場合 ( $\hat{p}_{kli} > \hat{P}_{k,U}$ ) は、修正倍率を調査区別上限値まで調整する ( $m'_{kli} = m_{kli,U}$ )。

なお、式(5)において、変更後の抽出率算出における端数の切上げ処理を省略し、 $w'_{kli,ALT} = h'_{kli}/20$ とした場合、式(6)のとおり準備調査時点の換算世帯数を用いることなく、換算世帯数が増加した場合の調査区別上限値 ( $m_{kli,U,ALT}$ ) が求められる。

$$m_{kli,U,ALT} = \frac{\hat{P}_{k,U} h_{kli} w'_{kli,ALT}}{g_{kl} p_{\text{国},kli} h'_{kli}} = \frac{\hat{P}_{k,U} h_{kli} h'_{kli}/20}{g_{kl} p_{\text{国},kli} h'_{kli}} = \frac{\hat{P}_{k,U} h_{kli}}{20 g_{kl} p_{\text{国},kli}} \dots (6)$$

式(6)より求めた調査区別上限値 ( $m_{kli,U,ALT}$ ) の全国及び地域別の分布は図3のとおりであった。地域別の平均値は概ね2に近い値となっており、分布の多くは概ね3より小さく、狭い範囲に収まっている。こうした特徴は平成22年国勢調査に固有のものではなく、平成27年国勢調査においても同じであった<sup>16</sup>。

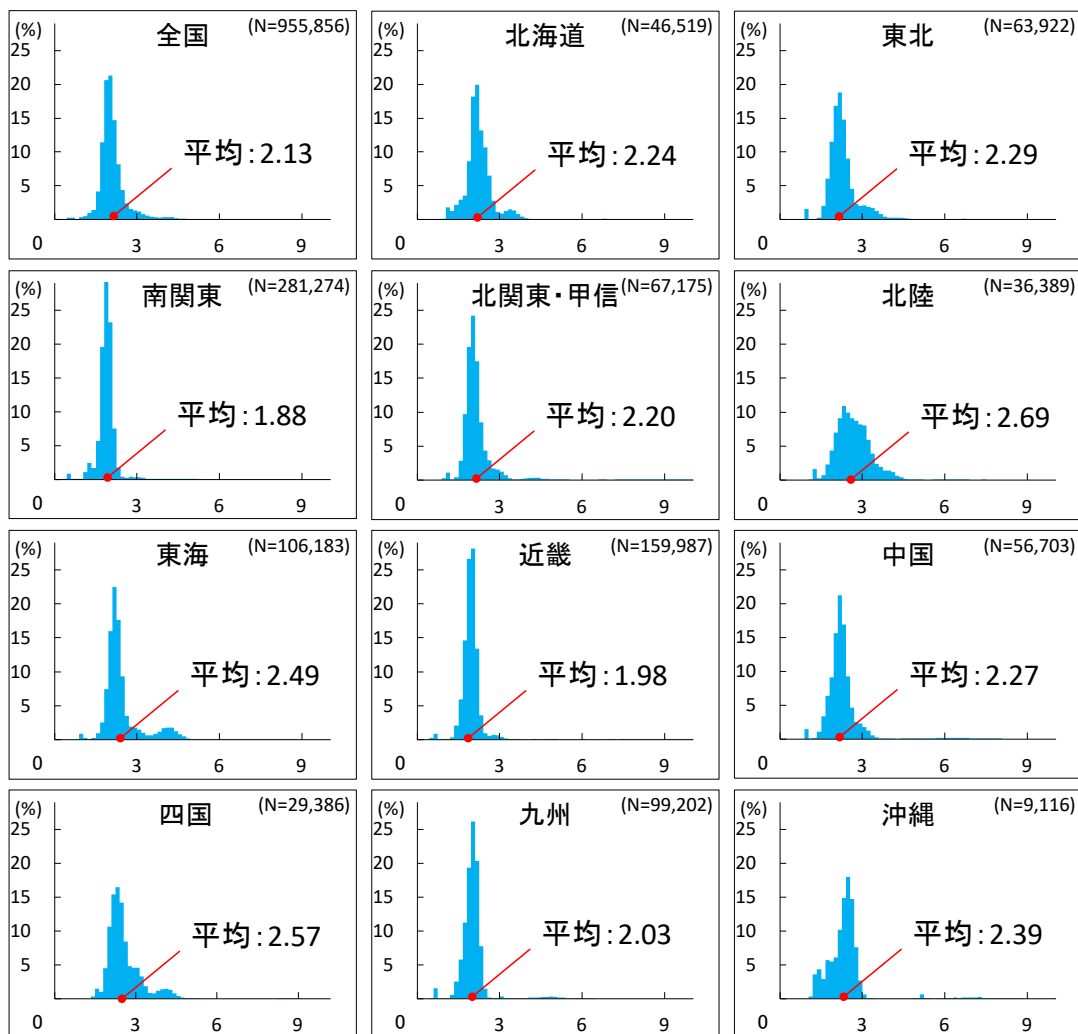


図3. 調査区別上限値 ( $m_{kli,U,ALT}$ ) 別調査区数割合

16 参考のため、平成27年国勢調査に基づく調査区別上限値 ( $m_{kli,U,ALT}$ ) 別調査区数割合を参考2に掲載する。

## 2 実績値を用いた調整結果の試算

### 2.1. 労働力調査実績（実際に住戸の抽出率を変更した調査区）

平成25年5月から31年4月までの72か月に労働力調査が開始された、平成22年国勢調査に基づく21,840調査区<sup>17</sup>のうち、調査区内の世帯数が国勢調査時点より増加したため抽出率を変更した調査区<sup>18</sup>及び分割した調査区（以下「抽出率変更調査区」という。）は表4のとおり、計456調査区であった。

そのうち、国勢調査時点で無人であった調査区は2調査区、2年目調査で抽出率を変更した調査区は10調査区であった。

表4. 1年目・2年目調査別抽出率変更調査区数

地域	平成25年5月から31年4月までに調査が開始された調査区				
		うち抽出率変更調査区			
		うち1年目抽出率変更 うち無人	うち2年目 抽出率変更		
全国	21,840	456	446	2	10
北海道	1,320	21	21	1	0
東北	1,740	32	30	0	2
南関東	4,500	147	146	1	1
北関東・甲信	1,800	34	33	0	1
北陸	1,320	15	15	0	0
東海	2,340	33	33	0	0
近畿	2,940	71	66	0	5
中国	1,500	20	20	0	0
四国	1,140	15	15	0	0
九州	2,160	23	22	0	1
沖縄	1,080	45	45	0	0

### 2.2. 調整対象の調査区

抽出率変更調査区について、準備調査時点における調整前の修正倍率 ( $m'_{kli}$ ) 及び単位推定人口 ( $\hat{p}'_{kli}$ )<sup>19</sup>を求め、これらが現行の調整方法である修正倍率2倍を超過する調査区及び地域別上限値 ( $\hat{P}_{k,U}$ ) を超過する調査区を確認した。

その結果、表5のとおり、修正倍率2倍を超過した調査区は全国で63調査区、地域別上限値 ( $\hat{P}_{k,U}$ ) を超過した調査区は54調査区と同程度の数であった。また、地域別に見てもそれぞれの基準を超過した調査区数はおおむね一致していた。抽出率変更調査区を層別に確認すると、表6のとおり、層22が150調査区と最も多く、そのうち22調査区が修正倍率2倍を超過し、19調査区が地域別上限値 ( $\hat{P}_{k,U}$ ) を超過していた。また、層別に見た場合でも、それぞれの基準を超過した調査区数はおおむね一致していた。

表5. 基準を超過した調査区数

地域	抽出率変更調査区			
		うち修正倍率2倍又は地域別上限値を超過した調査区		
		うち修正倍率2倍を超過した調査区	うち地域別上限値を超過した調査区	
全国	456	74	63	54
北海道	21	5	5	4
東北	32	6	6	5
南関東	147	30	24	24
北関東・甲信	34	2	2	2
北陸	15	1	1	0
東海	33	5	5	1
近畿	71	13	11	11
中国	20	2	1	1
四国	15	2	2	1
九州	23	3	3	2
沖縄	45	5	3	3

17 当該期間に調査を開始した平成22年国勢調査に基づく調査区について1年目調査と2年目調査を区別せず数えている。参考のため、当該期間の開始月別に集計したものを参考3に掲載する。  
 18 住戸が大幅に増加していない ( $h'_{kli} \leq 40$ ) にもかかわらず、前期・後期の調査を行うために住戸の抽出率の逆数を1から2に変更した調査区については対象外とした。  
 19 国勢調査時点で無人であるため、換算世帯数 ( $h_{kli}$ ) がゼロとなり式(3)により単位推定人口 ( $\hat{p}'_{kli}$ ) が求められない調査区は、2人以上一般世帯の平均世帯人員を3人とし、 $\hat{p}'_{kli} = 3g_{kl}m_{kli} h'_{kli}/w'_{kli}$  として求めた。

表6. 層別 抽出率変更調査区数

地域	( * )	計	層																							
			02	03	0401	0402	0403	0404	0422	0482	0493	05	11	12	13	14	15	16	17	18	19	20	21	22	23	99
全国	①	456	2	5	2	3	1	5	2	2	3	1	3	6	7	4	9	33	12	4	21	16	70	150	36	59
	②	63	2	5	1	1	1	4	1	1	1	1	1	1	1	1	1	1	2	1	4	2	3	22	4	7
	③	54	1	4	1	1	1	1	1	1	1	1	1	2	2	1	1	6	3	1	4	2	19	4	4	4
北海道	①	21	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	8	2	2
	②	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1
	③	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1
東北	①	32			2	2	1	1	1	1	1	1	1	1	1	1	1	3	1	1	1	2	5	9	2	3
	②	6			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	4	1	1	1
	③	5			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	4	1	1	1
南関東	①	147	1	2	1	1	1	1	1	1	1	1	3	4	4	1	1	2	3	3	15	2	16	59	17	20
	②	24	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	4	1	1	7	4	
	③	24	1	2	1	1	1	1	1	1	1	1	2	2	2	1	1	1	2	4	4	1	1	7	4	
北関東・甲信	①	34								1	1	1	1	1	2	2	1	1	1	1	1	5	11	6	1	5
	②	2																							2	
	③	2																							2	
北陸	①	15						1	1	1	1	1	1	1	1	1	1	4	2	1	1	1	2	1	1	1
	②	1						1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	③																									
東海	①	33		1					1								2	1	2	2	1	4	7	7	1	2
	②	5		1													1	1	1	1	1	1	1	1	1	
	③	1																								
近畿	①	71					1	1	1	1	1	1	1	1	1	1	6	6	1	1	1	18	25	3	12	
	②	11															1	2	2	2	2	2	6	2	2	
	③	11															1	2	2	2	2	5	5	1	1	
中国	①	20															2	2	2	1	2	6	2	1	4	
	②	1															1	1	1	1	1	1	1	1	1	
	③	1															1	1	1	1	1	1	1	1	1	
四国	①	15		1				1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	2	2	1	1
	②	2		1				1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	③	1		1				1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
九州	①	23							1							2	4	4	1	1	3	10	2	2	2	
	②	3															1	1	1	1	1	2	2	1	1	
	③	2															1	1	1	1	1	1	1	1	1	
沖縄	①	45						1	1	1	1	1	1	1	1	1	7	7	1	2	2	20	5	9	9	
	②	3						1	1	1	1	1	1	1	1	1	2	2	1	1	1	2	2	2	2	
	③	3															2	2	1	1	1	1	1	1	1	

( \* ) ①抽出率変更調査区 ②修正倍率2倍を超過した調査区 ③地域別上限値を超過した調査区

### 2.3. 調整結果の算出と比較

基準を超過した74調査区について、修正倍率 ( $m'_{kli}$ ) 調整前後の単位推定人口 ( $\hat{p}'_{kli}$ )<sup>20</sup>の最大値を確認したところ、表7のとおり北海道、東北、南関東、近畿、中国、九州及び沖縄の7地域及び全国で、修正倍率2倍を上限とする調整方法の方が大きく、特に、北海道、南関東及び全国については、調査区別上限値 ( $m_{kli,U}$ ) による調整方法の約1.5倍であった。一方、調整後の平均値は、東北、南関東、北関東・甲信、北陸、東海、近畿、四国、九州及び沖縄の9地域及び全国で、修正倍率2倍を上限とする調整方法が小さかった。

つまり、調査区別上限値 ( $m_{kli,U}$ ) による調整方法は、修正倍率2倍を上限とする調整方法と比較して平均的には推定値の調整幅が小さい、つまり弱い調整となっている一方で、地域別上限値 ( $\hat{P}_{k,U}$ ) を大きく超えるような単位推定人口 ( $\hat{p}'_{kli}$ ) については、強い調整を行っているといえる。

表7. 調整前後の単位推定人口 ( $\hat{p}'_{kli}$ )

地域	基準を超過した調査区	最大値 (人)			平均 (人)		
		修正倍率調整前	修正倍率調整後		修正倍率調整前	修正倍率調整後	
			修正倍率2倍 ( $m'_{kli} = 2$ )	調査区別上限値 ( $m'_{kli} = m_{kli,U}$ )		修正倍率2倍 ( $m'_{kli} = 2$ )	調査区別上限値 ( $m'_{kli} = m_{kli,U}$ )
全国	74	1,125,245	254,636	182,420	216,018	130,872	142,383
北海道	5	317,211	181,263	109,316	171,836	108,123	104,523
東北	6	405,571	162,228	142,256	201,801	115,692	133,489
南関東	30	1,125,245	254,636	172,876	277,933	159,886	167,881
北関東・甲信	2	283,066	133,190	139,835	230,337	119,703	139,835
北陸	1	54,870	36,580	54,870	54,870	36,580	54,870
東海	5	286,658	147,808	182,420	182,999	111,989	162,151
近畿	13	539,302	206,448	161,808	224,926	151,147	160,488
中国	2	153,512	153,512	132,300	123,828	117,109	113,222
四国	2	227,100	56,775	103,742	143,336	48,245	81,657
九州	3	205,049	153,767	142,639	172,523	125,942	136,339
沖縄	5	46,396	46,391	35,674	34,172	28,763	29,191

準備調査時点の単位推定人口 ( $\hat{p}'_{kli}$ ) と修正倍率 ( $m'_{kli}$ ) について調整前後の関係を確認すると、図4のとおり、地域別上限値 ( $\hat{P}_{k,U}$ ) を超過した調査区の修正倍率 ( $m'_{kli}$ ) を調査区別上限値 ( $m_{kli,U}$ ) に調整した場合、影響が大きな調査区のみを検出して調整していることが見て取れる。

20 国勢調査時点で無人であるため換算世帯数 ( $h_{kli}$ ) がゼロとなり、式(5)により調査区別上限値 ( $m_{kli,U}$ ) を求められない調査区については、式(6)により求めた調査区別上限値 ( $m_{kli,U,ALT}$ ) の全国平均値 ( $\bar{M}_{U,ALT}$ ) である、2.13を用いた。

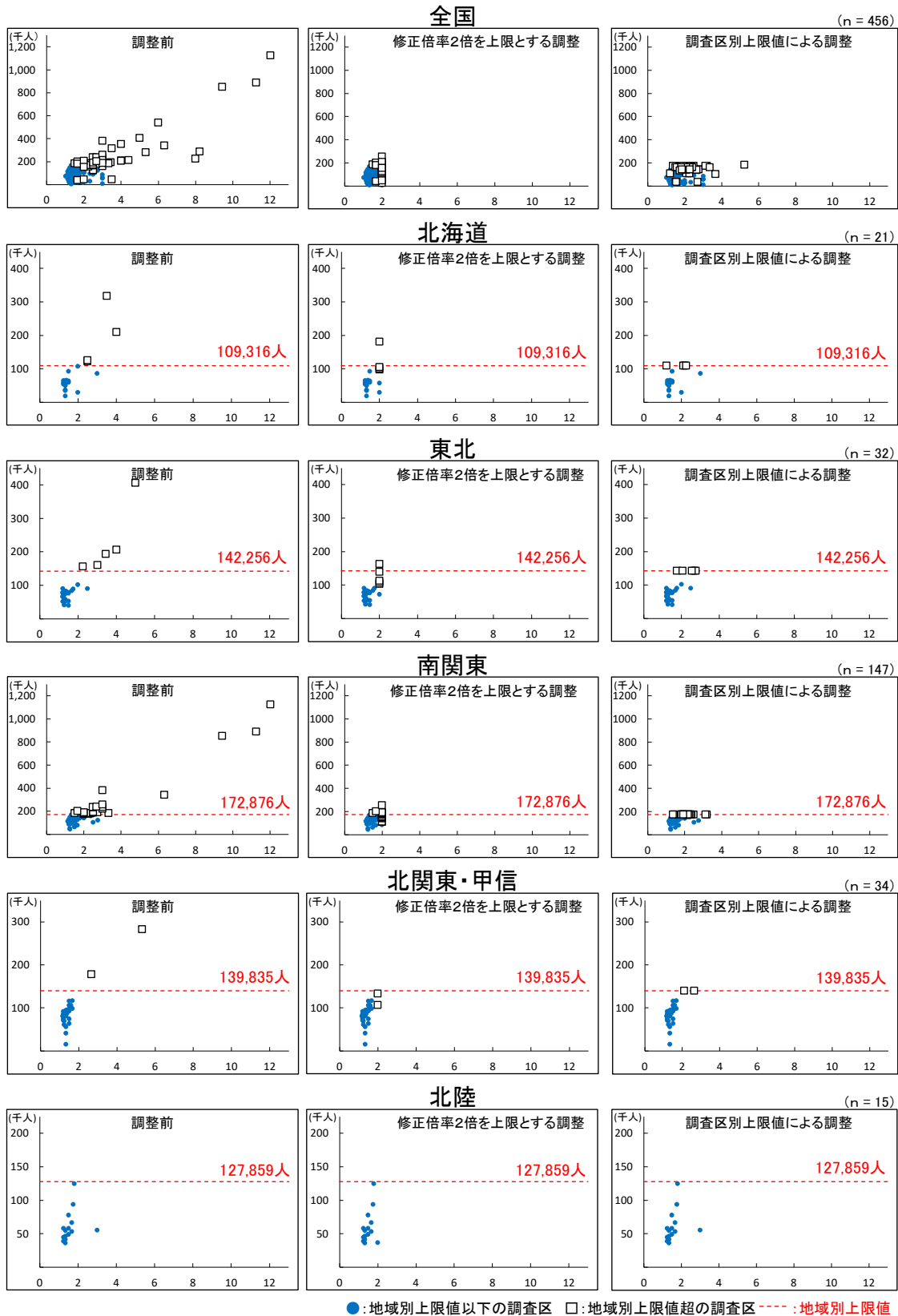


図4. 調整前後の修正倍率 ( $m'_{kli}$ ) 及び単位推定人口 ( $\hat{p}_{kli}$ )

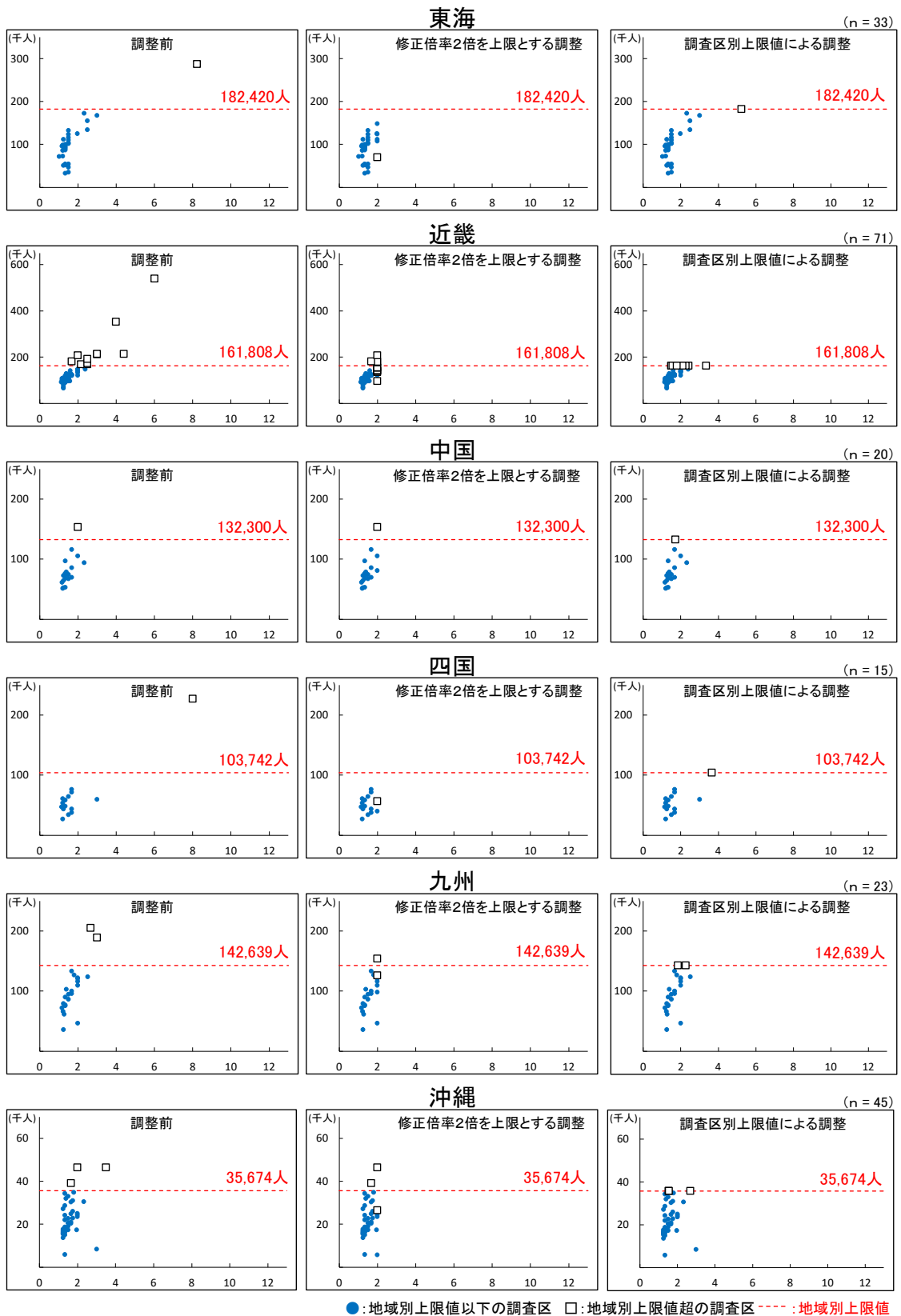


図4. 調整前後の修正倍率 ( $m'_{kli}$ ) 及び単位推定人口 ( $\hat{p}'_{kli}$ ) (つづき)



修正倍率2倍を上限とする調整方法においても、地域別上限値 ( $\hat{P}_{k,U}$ ) を超過した調査区は概ね調整されている。また、東海及び四国で、それぞれ1調査区が地域別上限値 ( $\hat{P}_{k,U}$ ) の約半分の大きさに調整されているものの、その他の地域では、調整後の値は地域別上限値 ( $\hat{P}_{k,U}$ ) を中心に分布している。なお、地域別上限値 ( $\hat{P}_{k,U}$ ) を下回る調査区も調整されているが、地域別上限値 ( $\hat{P}_{k,U}$ ) を超過した調査区と比べて、その調整幅は小さいものとなっている。地域別上限値 ( $\hat{P}_{k,U}$ ) を超過した調査区の調整方法による違いは特に南関東において顕著であり、調査区別上限値 ( $m_{kli,U}$ ) により調整する方法では一律に約17万人まで調整されている一方、修正倍率2倍を上限とする調整方法においては、約11万から25万人まで広い範囲に調整されている。

両調整方法による調整後の単位推定人口 ( $\hat{p}'_{kli}$ ) の関係を図5で確認すると、図4と同様に地域別上限値 ( $\hat{P}_{k,U}$ ) を超過した調査区において両調整方法の違いが顕著であることが見て取れる。

また、修正倍率2倍を上限とする調整方法では、地域別上限値 ( $\hat{P}_{k,U}$ ) を下回る調査区も調整されているものの、その調整幅は小さいことがここでも見て取れる。

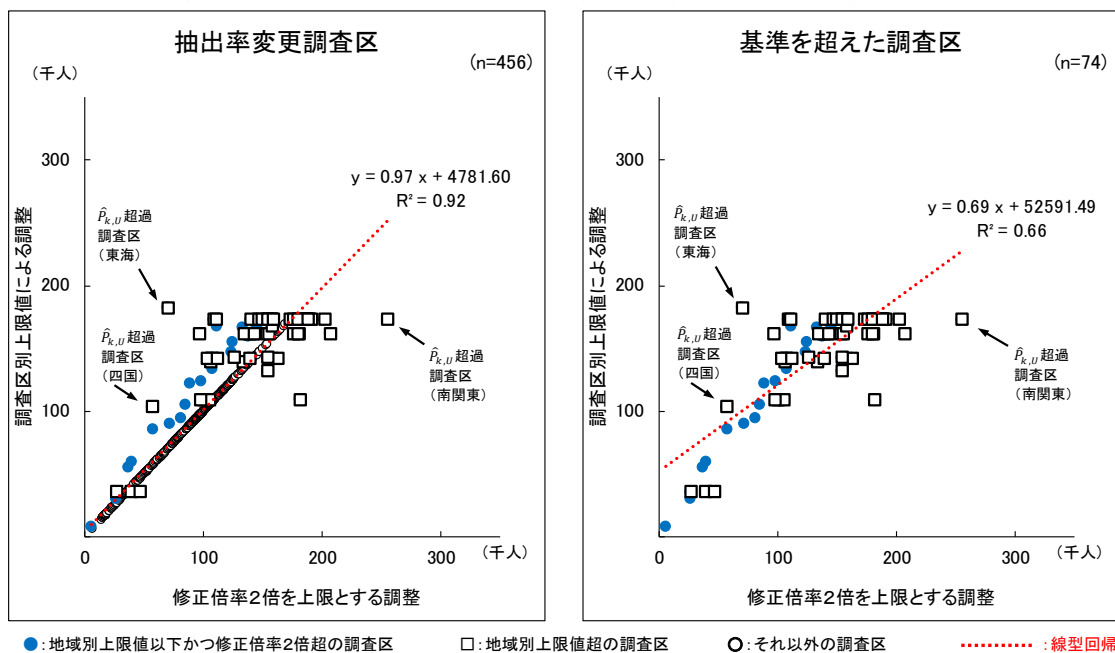


図5. 調整後の単位推定人口 ( $\hat{p}'_{kli}$ )

このように、調整後の単位推定人口 ( $\hat{p}'_{kli}$ ) を個別に比較した場合、特に南関東において両調整方法の特徴的な違いが確認できる。

ただし、前章で確認した調査区別上限値 ( $m_{kli,U,alt}$ ) の地域別平均値及び本章で確認した結果を総じて見ると、修正倍率2倍を上限とする調整は、調査区別上限値 ( $m_{kli,U}$ ) による調整に近いことが見て取れる。

### 3 まとめ

本稿では、母集団の推定値が真の値と大きく乖離する可能性への対応方法として、母集団である国勢調査の分布及び労働力調査の標本設計を吟味することにより、頑健な統計量といわれる四分位数と併せて、上位0.1%点を外れ値の基準に用いる以下の手法を試みた。

- ① 国勢調査の全調査区について単位推定人口 ( $\hat{p}_{\text{国},kli}$ ) を求め、その全国上位0.1%点と地域別四分位数を基に、労働力調査の外れ値の基準として地域別上限値 ( $\hat{P}_{k,U}$ ) を設定する
- ② 予め設定した地域別上限値 ( $\hat{P}_{k,U}$ ) により、推定結果に与える影響が極めて大きいと考えられる調査区のみを労働力調査の準備調査の結果から検出し、地域別上限値 ( $\hat{P}_{k,U}$ ) まで検出した調査区  $i$  の単位推定人口 ( $\hat{p}_{kli}$ ) を調整する

検証のため、労働力調査において住戸の抽出率を変更した調査区について、今回検討した調整方法と、修正倍率2倍を上限とする現行の調整方法とを比較した結果、今回検討した手法の方が、影響が大きな調査区のみを検出して、調整していることを確認した。また、現行の調整方法では影響が極めて大きいとは考えにくい調査区も調整されるため、平均的には今回検討した手法より強い調整となっていた。通常、標本から得られた調査結果は、母集団の姿を適切に反映していると考えられることから、今回検討した調整方法を採用することで、労働力調査の結果が短期的には安定し、長期的にはその精度が向上する可能性があると考えられる。こうした可能性について検証するため、国勢調査の結果及び人口推計の結果との比較等により、調整方法の違いが労働力調査の結果に与える影響を、推定値の安定性及びバイアスの点から分析することが望ましいと思われる。

調整の基準と手法の選定に当たっては、その基準と対応方法とが、ユーザーが理解しやすいシンプルなものとなることを心がけたが、そのようになれば幸いである。

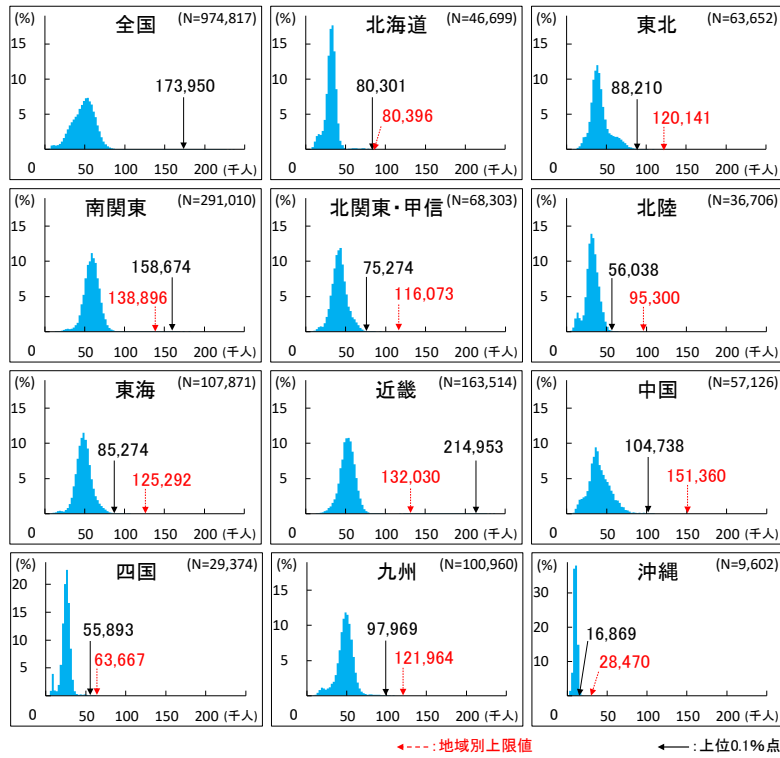
なお、本稿の目的とはしていないものの、実務においては統計作成の実務コストが大きく嵩まない調整方法が求められるが、今回検証した調整方法は調査区別上限値を算出する必要があること等からコストが嵩まないとは考えにくい。平成22年及び平成27年国勢調査から算出した調査区別上限値 ( $m_{kli,U,alt}$ ) の地域別平均値が概ね2倍となっていることなどから伺えるように、修正倍率の上限値を2倍とする現行の調整方法は、今回検証した調整方法の良い近似となっており、限られた時間の中で調査及び集計を行う必要のある労働力調査の実務に即した調整方法であると考えられる。

### 謝辞

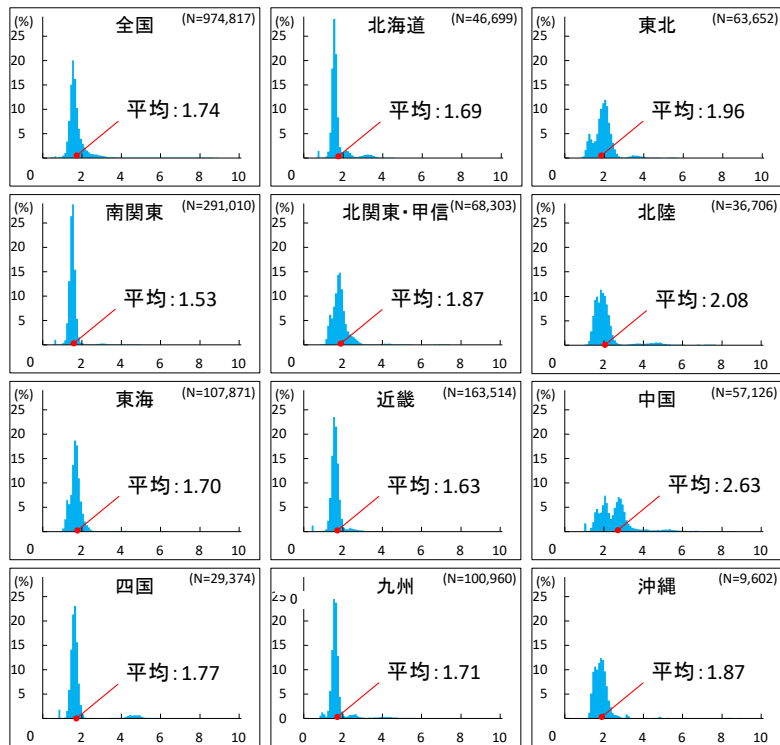
本研究の遂行にあたり、総務省統計局統計調査部国勢統計課労働力人口統計室及び独立行政法人統計センター情報技術センター技術研究開発課の職員の方々から有益なコメントをいただいた。感謝の意を表したい。また、本稿について丁寧な査読をしていただき、多くの改善点の指摘及び有益なコメントをしていただいた2名の匿名査読者にも感謝申し上げる。

### 参考文献

- [1] 石川 篤史・遠藤 峻介・白鳥 哲哉(2010)「ビジネスサーベイにおける外れ値対応」, 日本銀行ワーキングペーパーシリーズ, No.10-J-12
- [2] 野呂 竜夫・和田 かず美(2015)「統計実務におけるレンジチェックのための外れ値検出方法」, 統計研究彙報, 第72号
- [3] 総務省統計局(2013)労働力調査標本設計の解説, ISSN 2187-7882
- [4] 総務省統計局(2018)労働力調査標本設計の解説, ISSN 2187-7882



参考 1. 平成27年国勢調査に基づく単位推定人口 ( $\hat{p}_{\text{国},kli}$ ) 別調査区数割合、地域別上限値 ( $\hat{P}_{k,U}$ )、上位0.1%点



参考 2. 平成27年国勢調査に基づく調査区別上限値 ( $m_{kli,U,ALT}$ ) 別調査区数割合

参考3. 1年目・2年目調査別抽出率変更調査区数、調整前修正倍率 ( $m'_{kli}$ ) <sup>21</sup>

平成25年度

		計	5月	6月	7月	8月	9月	10月	11月	12月	1月	2月	3月	4月	
1年目	調査区数	75	10	5	11	5	6	8	4	6	7	7	2	4	
	修正倍率	平均	1.52	1.35	2.20	1.49	1.80	1.56	1.42	1.31	1.47	1.69	1.37	1.46	1.33
		最大値	3.50	1.67	3.50	2.00	2.75	2.00	1.67	1.33	2.00	2.50	1.75	1.59	1.50
		最小値	1.17	1.20	1.33	1.25	1.25	1.25	1.33	1.25	1.17	1.25	1.17	1.33	1.25

平成26年度

		計	5月	6月	7月	8月	9月	10月	11月	12月	1月	2月	3月	4月	
1年目	調査区数	61	6	6	4	4	5	4	9	5	2	4	7	5	
	修正倍率	平均	1.70	1.37	1.53	1.38	2.13	1.72	2.13	2.25	1.67	1.27	1.30	1.58	1.59
		最大値	6.32	1.50	2.00	1.67	4.00	2.31	3.50	6.32	3.00	1.33	1.33	2.50	1.80
		最小値	1.02	1.20	1.33	1.25	1.33	1.50	1.17	1.02	1.20	1.20	1.20	1.17	1.33
2年目	調査区数	2									1				
	修正倍率	平均	2.50					2.00				3.00			
		最大値	3.00					2.00				3.00			
		最小値	2.00					2.00				3.00			

平成27年度

		計	5月	6月	7月	8月	9月	10月	11月	12月	1月	2月	3月	4月	
1年目	調査区数	89	5	7	4	3	10	9	10	7	8	9	10	7	
	修正倍率	平均	1.91	1.65	2.00	1.52	1.33	1.46	3.95	1.65	1.48	1.54	1.62	2.09	1.88
		最大値	12.03	2.00	3.01	2.00	1.50	2.50	12.03	2.50	2.41	2.00	3.00	3.45	3.00
		最小値	1.14	1.25	1.33	1.25	1.25	1.14	1.20	1.25	1.20	1.25	1.25	1.25	1.20
2年目	調査区数	2	1										1		
	修正倍率	平均	5.67	6.00									5.33		
		最大値	6.00	6.00									5.33		
		最小値	5.33	6.00									5.33		

平成28年度

		計	5月	6月	7月	8月	9月	10月	11月	12月	1月	2月	3月	4月	
1年目	調査区数	115	4	4	16	9	13	9	10	8	14	8	11	9	
	修正倍率	平均	1.71	1.74	2.08	1.51	1.40	1.77	2.33	1.56	1.54	1.49	1.70	2.31	1.45
		最大値	9.44	3.00	2.50	2.50	2.00	4.41	8.23	2.25	2.33	3.00	2.67	9.44	2.00
		最小値	1.15	1.20	1.33	1.20	1.20	1.25	1.20	1.20	1.25	1.15	1.20	1.25	1.20
2年目	調査区数	4				3								1	
	修正倍率	平均	3.17				2.56							5.00	
		最大値	5.00				4.00							5.00	
		最小値	1.67				1.67							5.00	

平成29年度

		計	5月	6月	7月	8月	9月	10月	11月	12月	1月	2月	3月	4月	
1年目	調査区数	106	7	6	8	13	9	8	9	11	11	5	10	9	
	修正倍率	平均	1.62	1.46	1.37	1.72	1.50	1.55	1.68	1.53	1.50	2.19	1.80	1.51	1.61
		最大値	8.00	2.00	1.67	2.50	2.33	2.00	2.33	3.00	2.00	8.00	2.67	2.50	2.73
		最小値	1.14	1.25	1.25	1.33	1.17	1.23	1.33	1.20	1.25	1.14	1.20	1.25	1.20
2年目	調査区数	0													
	修正倍率	平均	-												
		最大値	-												
		最小値	-												

平成30年度

		計	5月	6月	7月	8月	9月	10月	11月	12月	1月	2月	3月	4月	
2年目	調査区数	2						1						1	
	修正倍率	平均	2.00					2.00						2.00	
		最大値	2.00					2.00						2.00	
		最小値	2.00					2.00						2.00	

21 平成22年国勢調査に基づく調査区のうち、国勢調査以降に調査区内の世帯数が増加したため抽出率を変更した調査区及び分割した調査区を集計対象としている。また、1年目調査時点で変更した抽出率を2年目調査に用いた調査区は、2年目の集計対象から除外している。