

国勢調査における匿名データの作成とその検証

**NSTAC**

---

*Working Paper No.27*

平成 26 年 10 月

独立行政法人 統計センター

製表技術参考資料は、独立行政法人 統計センターの職員がその業務に関連して行った製表技術に関する研究の結果を紹介するためのものである。

ただし、本資料に示された見解は、執筆者の個人的見解である。

## 目 次

要旨	1
1 はじめに	3
2 欧米諸国における人口センサスの匿名化マイクロデータの提供と匿名化措置	3
(1) アメリカ	3
(2) カナダ	4
(3) イギリス	5
3 国勢調査の匿名データの特徴	5
4 国勢調査の匿名データの作成に用いられた匿名化措置の概要	6
(1) 情報の削除(suppression)	6
(2) 分類区分の再編(リコーディング(recoding))	8
(3) トップ(ボトム)・コーディング(top(bottom)coding)	8
(4) サンプリング(sampling)	8
(5) 外部参照情報に基づくレコードの削除ー公表済みの結果表に着目してー	10
(6) スワッピング(data swapping)	10
5 国勢調査の匿名データの作成に関する検証	11
(1) 情報の削除に関する検証	12
(2) サンプリングに関する検証	13
(3) 度数1または2に該当するレコード削除に関する検証	13
(4) スワッピングに関する検証	16
6 結びにかえて	16
参考文献	18
付表1 原データと1%抽出データの分布の比較	19



## 国勢調査における匿名データの作成とその検証

星野なおみ\*、伊藤伸介\*\*、後藤武彦\*\*\*

### 要旨

我が国では、2013年12月末に平成12年国勢調査の匿名データが、2014年3月末には平成17年国勢調査の匿名データが、それぞれ提供されることとなった。国勢調査の匿名データの作成にあたって、(独)統計センター統計情報・技術部統計技術研究課は、総務省統計局との共同研究という形で、国勢調査の匿名データの作成方法に関する研究を進めてきた。本稿は、統計技術研究課でこれまで行ってきた国勢調査の匿名データの作成方法に関する基礎研究の概要だけでなく、国勢調査の匿名データの作成に関する具体的な検証内容について論じている。

国勢調査の匿名データの作成においては、レコード削除、サンプリング、年齢等のリコーディングやトップ(ボトム)・コーディングが匿名化手法として適用されている。国勢調査の匿名データにおいてレコードの削除を行う場合には、①0.5%基準に基づき、母集団全体に占める比率が0.5%以下となるセルに該当するレコードの削除、および②特異なセルの分布状況を確認し、その分布から設定された基準をもとに選別されたレコードの削除が行われる。さらに、公表されている外観識別性の高い属性間の結果表が、匿名データに対する外部参照情報となりうることから、外観識別性の高い属性間の結果表において度数1や2を含むセルに該当するレコードの削除も行われている。これらの匿名化手法が非攪乱的手法であるのに対して、国勢調査の匿名データにおける秘匿性を確保するために、一部のレコードに対して攪乱的手法としてスワッピングが適用されている。

こういった国勢調査の匿名データの作成にあたって、本研究では、使用した匿名化手法ごとに、有用性と秘匿性に関する検証を行った。具体的には、レコードの削除については、外観識別性の高い属性を含めたクロス表を作成し、クロス表における度数1および2の出現状況を確認した上で、出現した度数1や2における特異な分布の確認を行った。また、サンプリングに関しては、世帯単位と個人単位のそれぞれでレコード群の一部を抽出した場合の結果の比較を行った。さらに、結果表における度数1や2の削除に関しては、外観識別性の高い属性を用いてテストデータを作成した上で、擬似的なクロス表を作成し、度数1または2に該当する個人が含まれる世帯を削除した場合の分布状況を確認した。スワッピングについては、スワッピング率を変えた場合のスワッピングの有効性に関する検証

\* (独) 統計センター統計情報・技術部統計技術研究課

\*\* (独) 統計センター統計情報・技術部統計技術研究課非常勤研究員 (中央大学経済学部准教授)

\*\*\* (独) 統計センター統計情報・技術部共同利用システム課(元統計技術研究課)

を行った。本研究は、統計実務の観点から、これまで統計センターで行われてきた匿名化技法に関する研究の適用可能性を追究した点で、有意義だと思われる。

## 国勢調査における匿名データの作成とその検証

星野なおみ、伊藤伸介、後藤武彦

### 1. はじめに

2009年4月に統計法が改正されて以降、公的統計(政府統計)の匿名データの提供が可能になった。総務省統計局では、住宅・土地統計調査、全国消費実態調査、就業構造基本調査、社会生活基本調査といった標本調査の匿名データが、これまで提供されており、就業やライフスタイルの分野を中心に、ミクロレベルの実証的な社会経済研究に寄与してきた。一方、我が国においても、全数調査である国勢調査の匿名データの提供に対するニーズが高かったことから、2013年12月末に平成12年国勢調査の匿名データ、2014年3月末には平成17年国勢調査の匿名データがそれぞれ作成・提供されることとなった。

ところで、(独)統計センター統計情報・技術部統計技術研究課は、様々な匿名化技法を適用することによって作成されたマイクロデータ(以下「匿名化マイクロデータ」と呼ぶ。)のさらなる展開可能性を目指して、匿名化技法の有効性、および秘匿性と有用性の評価に関する実証研究をこれまで進めてきたが(伊藤他(2008)、伊藤他(2009))、伊藤他(2010)、伊藤・村田(2013)、伊藤・星野(2013))、今回、総務省統計局との共同研究という形で、国勢調査の匿名データの作成の実務に関わる研究を行ってきた。本稿は、国勢調査の匿名データの作成方法に関する基本的な特徴を述べるだけでなく、統計技術研究課において進めてきた国勢調査の匿名データの作成に関する実証研究の成果についてまとめたものである。

### 2. 欧米諸国における人口センサスの匿名化マイクロデータの提供と匿名化措置

1960年代以降、アメリカ、カナダ、イギリスといった欧米諸国において、政府統計マイクロデータの提供が進められてきた。とりわけ、人口センサスの個票データに対しては、非攪乱的手法(non-perturbative methods)だけでなく、攪乱的手法(perturbation)を施すことによって、匿名化マイクロデータが作成・提供されてきた。本節では、アメリカ、カナダ、イギリスを例に、人口センサスにおける匿名化マイクロデータの提供状況を述べることにしたい。

#### (1) アメリカ

アメリカでは、1963年より人口センサスのマイクロデータの提供が行われているが、2000年人口センサスの詳細調査票(long form)については、標本抽出率が1%と5%の2種類の一般公開型マイクロデータ(Public Use Microdata Samples=PUMS)が利用可能になっている(Zayatz(2007))。アメリカセンサス局は、一般公開型マイクロデータを作成するための非攪乱的な手法として、地理的な関

値の設定(Geographic Thresholds)、カテゴリカルな属性における分類区分の閾値(Categorical Thresholds)、トップ・コーディングを用いている。例えば、地理的な閾値については、2000年人口センサスの5%PUMSの場合、PUMSで利用可能な地域区分(Public Use Microdata Area)として、すべての地域が最低10万人の人口を有するような地域区分が設定されている。また、カテゴリカルな属性における分類区分の閾値に関しては、すべてのカテゴリカルな変数を対象に、公表された分類区分のおのおのについて全国レベルで少なくとも10,000人以上存在するように、分類区分が設定される。さらに、half-percent/three percent ruleに基づいてトップコーディングも適用されている<sup>1</sup>。一方、アメリカセンサス局は、攪乱的な手法としてラウンディング(伝統的な丸め, traditional rounding)、ノイズの付加(noise addition)、スワッピング(data swapping)を用いている。例えば、年齢におけるノイズの付加については、世帯人員が10人以上のような識別されるリスクが高い世帯を対象に、世帯員の年齢にノイズが導入されている(Zayatz(2007))。

## (2) カナダ

カナダでは、1971年人口センサスから、PUMF(=Public Use Microdata Files)が作成・提供されてきた(赤谷・荒川・伊藤(2014))。カナダ統計局は、1971年から2001年までの人口センサスについては、個人ファイル(Individuals file)、家族ファイル(Families file)及び世帯・住宅ファイル(Households and Housing file)と呼ばれる3種類のPUMFを作成してきた。一方で、2006年の人口センサスのPUMFにおいては、標本抽出率が約2.7%で数多くの社会変数が含まれる個人ファイルと、標本抽出率が約1%で地域区分も粗いが、同一の世帯及び家族内の構成員と連結することが可能な階層ファイル(Hierarchical file)の2種類が提供されている。

人口センサスのPUMFの作成においては、リコーディング、トップコーディングといった非攪乱的な手法(non-perturbative methods)が主な匿名化技法として用いられているが、ラウンディング(丸め)といった方法も適用されている。リコーディングが用いられている主な社会変数は、年齢や産業であって、例えば、年齢は5歳階級区分で提供されている。また、トップコーディングを行っている変数は、主として、収入、住居費と部屋の数であり、例えば収入に関しては、地域や性別ごとに、上位1パーセントがトップコーディングの閾値に設定される。なお、トップコーディングが適用される変数の値については、トップコーディングの対象となるレコードにおける平均値が用いられている。一方、人口センサスのPUMFにおいて、特異なレコードが数多く存在する場合には、特異なレコードの数が許容可能な水準に達するまで、変数の分類区分の統合が行われる。さらに、特異なレコード以外にセンシティブな変数が含まれていた場合には、変数の削除等の処理が行われる(二次的な露見に対する変数削除(residual disclosure suppression))<sup>2</sup>。

<sup>1</sup> half-percent/three percent ruleの概要は以下のとおりである。

①全体の集団を対象とする属性(ex. 年齢)については、その属性の最上位となる階級(top code)が、全レコードの少なくとも0.5%を含んでいなければならない。  
 ②部分集団を対象とする属性(ex. 農業収入)については、その属性の最上位となる階級(top code)は、非負の属性値をもつレコード群の少なくとも3%か、あるいは全レコードの少なくとも0.5%を含んでいなければならない。その場合、より高い数値がトップコーディングの閾値として用いられる。

<sup>2</sup> カナダ統計局は、人口センサスのPUMFにおける秘匿性を検証するために、一意性の分析(unique analysis)を行

### (3) イギリス

イギリスでは、1991年人口センサスから、匿名化標本データ(Samples of Anonymised Records=SARs)の提供が開始された(森(2000, 51頁))。1991年人口センサスについては、世帯単位で抽出され、詳細な社会経済的属性を有する世帯 SAR と(Household SAR, 標本抽出率は1%), 個人単位で抽出され、世帯 SAR よりも地域区分が詳細な個人 SAR(Individual SAR, 標本抽出率は2%)が作成・提供されている。それに対して、2001年人口センサスでは、Special License(SL)型の世帯 SAR と標本抽出率が3%である個人 SAR(Individual SAR(Licensed))が提供されている。また、2001年個人 SAR については、1991年における区分と比較して地域区分が粗いことから、小地域分析用のマイクロデータ提供への利用者側のニーズが高くなった。そのため、2001年の人口センサスでは、小地域マイクロデータ(Small Area Microdata=SAM)が作成されている(伊藤(2011))。なお、イギリス国家統計局(The Office for National Statistics=ONS)は、なお、2011年人口センサスに関して、2014年にマイクロデータを作成・提供することが順次予定されているが、1%抽出の教育用個人ファイル(1 per cent Teaching individual file)は、すでに一般公開用のマイクロデータとして ONS から提供されている(伊藤(2014))。

1991年の世帯 SAR と個人 SAR の作成に用いられている匿名化技法は、(1)レコードの標本抽出、(2)データの削除(suppression)、(3)地域特性の制限、(4)個人・世帯属性の分類区分の統合であって(Marsh(1991), Marsh *et al.*(1994), 森(2000)等)、非攪乱的手法のみが用いられている。他方、2001年 SARs の作成においては、特殊な一意(special uniques)に該当するレコードに対して秘匿処理を行うために、上記の非攪乱的手法に加えて、攪乱的手法として、PRAM(Post Randomization Method)が用いられている(De Kort and Wathan(2009))。

## 3. 国勢調査の匿名データの特徴

国勢調査は、人口、世帯、産業構造等の実態を明らかにすることを目的とした、最も基本的な統計調査である。国勢調査は、10年ごとに実施される大規模調査とその中間年に行われる簡易調査に大別される。今回提供する匿名データでは、大規模調査の平成12年と簡易調査の平成17年を提供することとした。2か年分の匿名データが提供されることによって、2時点間の比較が可能になる。大規模調査と簡易調査に共通する調査項目については、性別、続柄、国籍、従業上の地位、世帯人員数、世帯の種類や住宅の建て方などがある。その一方、大規模調査のみの調査項目については、教育や家計の収入の種類などが該当する。

表1は、匿名データで利用可能な調査事項と適用される匿名化措置の一覧を示したものである。国勢調査の匿名データは「匿名データの作成・提供に関するガイドライン」に沿って作成されている。最初に、匿名データの作成にあたり、秘匿処理前の個票データ(以下「原データ」と呼ぶ。)か

---

っている。一意性の分析においては、様々な3変数についてのクロス集計表を作成し、特異なレコードが存在するかどうかを探索的に確認する。また、PUMFの提供については、露見リスク(disclosure risk)をゼロにするのではなく、露見リスクの制御(disclosure control)という考え方に基づいて、秘匿処理を行っている(赤谷・荒川・伊藤(2014, 4~5頁))。

らレコードの一部を削除した上で、世帯単位<sup>3</sup>で1%抽出を行っている(平成12年国調の場合には、約124万レコード)<sup>4</sup>。地域については、詳細な地域区分に対するニーズを考慮して都道府県と人口50万以上市区が提供されている。さらに、外観識別性の高い属性を対象に、リコーディングやトップ(ボトム)・コーディングを適用している。具体的には、世帯の家族類型、年齢、国籍、世帯主との続き柄、産業、職業等の調査事項においてリコーディングを行っており、年齢や就業時間においてトップ・コーディングを適用している。さらに、国勢調査の匿名データについては、秘匿性を確保するために、一部のレコードに対してスワッピングを適用しているのが、特徴的である。

#### 4. 国勢調査の匿名データの作成に用いられた匿名化措置の概要

本節では、国勢調査の匿名データの作成に用いられた匿名化技法の考え方について具体的に見ていくことにしたい。

##### (1) 情報の削除 (suppression)

情報の削除については、秘匿性の程度を高めるために、特異なレコードや特定化に用いられる変数が削除の対象となる。具体的には、国勢調査の匿名データの作成においては、母集団を対象に外観識別性の高い属性について度数分布をとった場合に、母集団全体に占める比率が0.5%以下に該当するレコードは、削除対象となるレコードと考えられている。

ところで、世帯人員の分布は地域によってばらつきがあるため、全国一律に0.5%の基準を適用した場合、多くのレコードが削除されることから、有用性が相対的に低くなる可能性がある。そのため、地域ごとに0.5%以下に該当する分類区分を確認した上で、地域ごとにレコード削除のための閾値を設定することで、削除の対象となるレコード数を少なくすることが可能になる。

図1は、地域別の世帯人員の度数分布を示したものであるが、0.5%以下に該当する区分はそれぞれ、A地域では9人世帯と10人以上世帯の2区分、B地域では8人世帯と9人世帯、10人以上世帯の3区分であって、A地域とB地域では、レコード削除の閾値が異なる。この閾値に沿った形で該当するレコードが削除される。

<sup>3</sup> 抽出単位については、一般世帯が世帯単位で抽出されているのに対して、施設等の世帯は、個人単位で抽出されている。施設等の世帯は施設を1つの世帯とみなして調査しているため、病院であれば1床を1人とみなして抽出している。

<sup>4</sup> 先述のとおり、アメリカにおける人口センサスのPUMS、カナダの人口センサスPUMFの階層ファイル、イギリスにおける人口センサスの世帯SAR(Special License(SL)型の世帯SAR)における標本抽出率はいずれも1%である。

表1 国勢調査の匿名データにおける提供項目と適用される匿名化措置の一覧

調査項目	平成12年	平成17年	匿名化措置
調査年月	○	○	
市区町村コード			
都道府県	○	○	
市区町村	●	●	・人口50万以上の市区のみ提供
調査区番号	—	—	
世帯番号・調査区内連番	—	—	
世帯員番号	○	○	
世帯の種類	●	●	・「一般世帯」及び「施設等の世帯」の2区分を提供 ・「施設等の世帯」の内訳は提供しない
世帯人員	●	●	・「一般世帯」については、地域区分により、世帯人員が7～9人以上いる世帯を削除 ・「施設等の世帯」については提供しない
世帯の家族類型	●	●	・6区分による提供
(再掲)3世代世帯	○	○	
家計の収入の種類	—	—	
住居の種類・住宅所有の関係	●	●	・「公営の借家」及び「都市機構・公社の借家」を統合 ・「給与住宅」及び「間借り」を統合 ・「住宅以外に住む一般世帯」の内訳は提供しない
住宅の床面積	●	●	・実数ではなく、階級で提供 ・「200～249㎡」及び「250㎡以上」を統合
住宅の建て方	●	●	・「長屋建」及び「その他」を統合
共同住宅の階数			
建物全体の階数	●	●	・実数ではなく、階級で提供 ・高層階の階数区分(「6～10階」、「11～14階」及び「15階以上」)を地域区分によって「6階以上」又は「11階以上」に統合
世帯が住んでいる階	●	●	・実数ではなく、階級で提供 ・高層階の階数区分(「3～5階」、「6～10階」、「11～14階」及び「15階以上」)を地域区分によって「3階以上」、「6階以上」又は「11階以上」に統合
出生の年月			
元号	—	—	
年	—	—	・出生年月は年齢に置き換える ・5歳階級で提供
月	—	—	・85歳以上をトップコーディング
年齢	●	●	
世帯主との続き柄	●	●	・「他の親族」、「住み込みの雇い人」及び「その他」を統合
男女	○	○	
配偶関係	○	○	
国籍	●	●	・「日本人」及び「外国人」の2区分とし、外国籍の内訳は提供しない
現在の住居における住居期間	○	○	
5年前の住居の所在地	●	○	・「他県から」及び「国外から」を統合
在学か否かの別	○	○	
学校の種類・未就学の種類	○	○	
労働力状態	●	●	・「家事などのほか仕事」及び「通学のかたわら仕事」を統合
就業時間			
実数	●	●	・90時間以上をトップコーディング
14区分	○	○	
10区分	○	○	
(再掲)40時間以下	○	○	
(再掲)42時間以下	○	○	
(再掲)48時間以下	○	○	
従業上の地位	●	●	・「雇人のある業主」、「雇人のない業主」及び「家庭内職者」を統合
産業(大分類)	●	●	・「農業」、「林業」及び「漁業」を統合 ・「鉱業」及び「建設業」を統合 ・「製造業」及び「電気・ガス・熱供給・水道業」を統合 ・「複合サービス業」及び「サービス業(他に分類されないもの)」を統合(平成17年のみ)
職業(大分類)	●	●	・「保安職業従事者」、「農林漁業作業員」及び「運輸・通信従事者」を統合
常住地による従業地・通学地	●	●	・「県内他市区町村で従業・通学」及び「他県で従業・通学」を統合
利用交通手段	●	○	・「利用交通手段が1種類」について、「勤め先・学校のバス」、「ハイヤー・タクシー」、「オートバイ」及び「その他」を統合

- そのまま提供
- 匿名化措置を講じて提供
- 提供しない

図1 情報の削除のイメージ図

	世帯人員					
	1人	...	7人	8人	9人	10人以上
A地域	25.4	...	1.8	0.53	0.12	0.04
B地域	27.8	...	1.4	0.46	0.11	0.02

(2) 分類区分の再編 (リコーディング (recoding))

分類区分の再編の1つであるリコーディング (区分の統合) は、類似した区分同士をひとつにまとめて、区分を粗くすることから、本研究では、情報の削除と同様に、母集団を対象に外観識別性の高い属性について度数分布をとった場合に、分類区分ごとの母集団全体に占める比率が0.5%以下となる区分を確認している。0.5%以下の区分を含む属性については、他の属性とのクロス表を作成した場合に、度数1または2が多く出現する可能性が高い。図2は、国籍を例にした分類区分の再編 (区分の統合) のイメージ図である。図2で示されるように、国籍では、日本人以外の区分に0.5%以下が多く存在することがわかる。そこで、他の属性とのクロス表を作成した場合、日本人以外の区分で度数1または2が多く出現することが明らかになった。そこで、日本人以外の区分を統合し、「外国人」という区分にしている。このように、国勢調査の匿名データの場合、有用性の観点から類似した区分同士を統合するだけでなく、0.5%以下に該当する区分をまとめることによって、秘匿性を確保している。

(3) トップ(ボトム)・コーディング (top(bottom)coding)

本研究においては、ある上限値以上(あるいは下限値以下)の区分を統合するトップ(ボトム)・コーディングについても、レコード削除やリコーディングと同様に、分類区分ごとの母集団全体に占める比率について0.5%以下となる区分を確認する。場合によっては、他の属性とのクロス表の分布状況も考慮する。例えば、図3は、年齢5歳階級に関するトップ・コーディングのイメージ図を示したものである。

年齢5歳階級では、90歳以上の分類区分は全て0.5%以下となる区分に該当しており、それらの区分を合計しても比率は、0.5%以下と算定される。そこで、85歳以上をひとつの区分にまとめることで0.5%以上となり、秘匿性を確保している。

(4) サンプリング (sampling)

国勢調査の匿名データの作成のために、サンプリングが行われる。図4は、母集団と標本データの関係を示したものであるが、標本データに「80歳女性役員」に該当するレコードが一意に存在したとしても(標本一意(sample unique))、母集団には、「80歳女性役員」に

図2 分類区分の再編 (区分の統合) のイメージ図

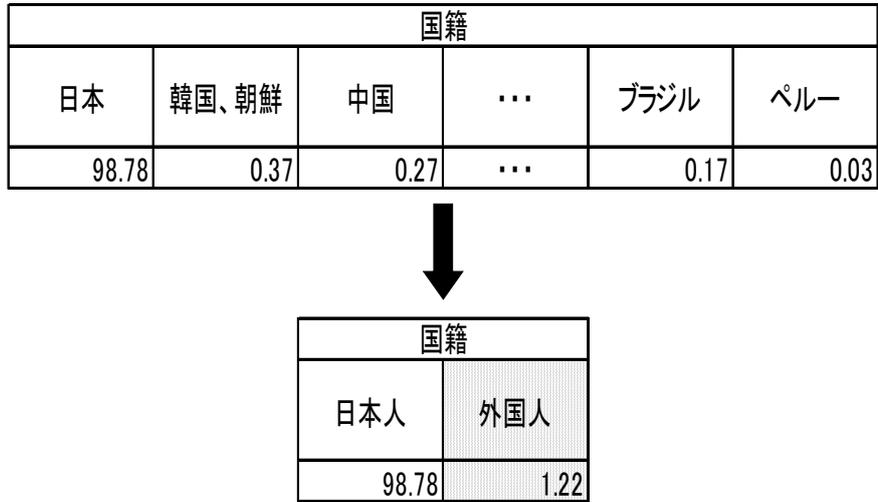


図3 トップ・コーディングのイメージ図

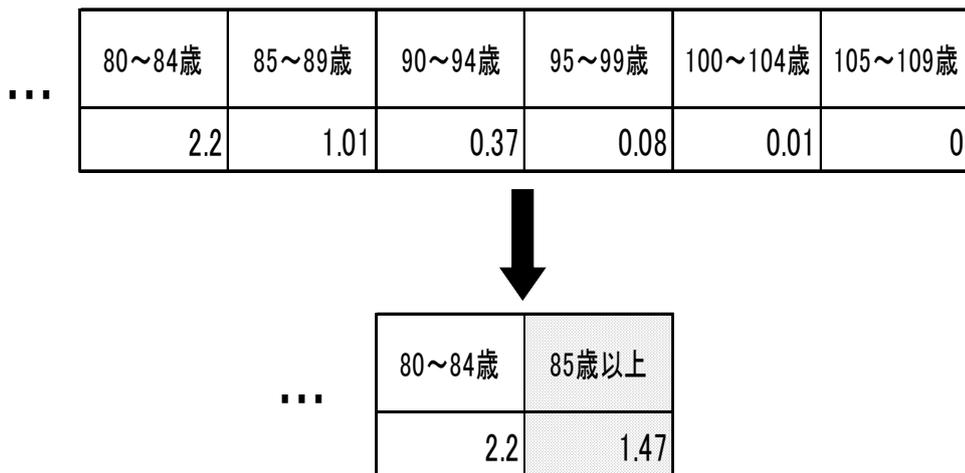
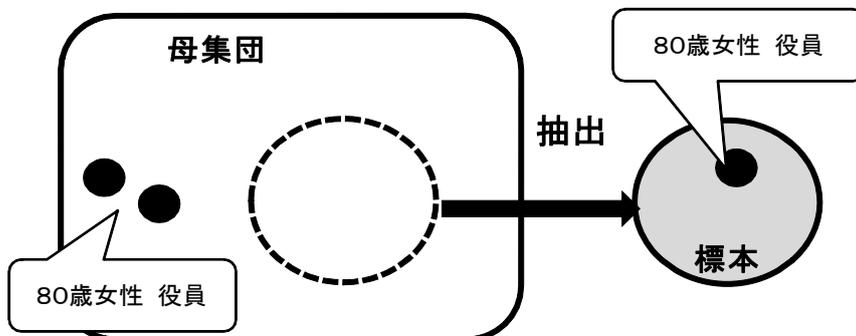


図4 サンプリングのイメージ図



該当する複数のレコードが含まれており、母集団一意(population unique)でない可能性も考えられる。このように、標本データに一意となる個体が存在しても、母集団一意でないことから、特定される可能性は低い。

国勢調査の匿名データの作成においては、世帯人員が多い世帯といった特異なレコードの削除を行った上で、世帯の種類ごとに、市区町村番号及び世帯人員で並べ替え、一般世帯については世帯単位、施設等の世帯については個人単位でそれぞれ系統抽出を行っている。なお、全国一律で1%の抽出となることから、匿名データに含まれるレコードにおいては、乗率(ウェイト)は付与されていない。100倍することで公表値に近い数値となるが、世帯人員が多い世帯などを削除しているため、公表値より若干小さい値となる。

#### (5) 外部参照情報に基づくレコードの削除—公表済みの結果表に着目して—

国勢調査の公表済み結果表には、外観識別性の高い属性間の結果表も存在する。そうした結果表に度数1や2を含むセルが存在した場合、そのようなセルと匿名データとつぎ合わせることによって、公表済み結果表が強力な外部情報となることが考えられる。例えば、性別と国籍の公表済み結果表内に度数1が存在した場合、これは母集団一意に該当するため、公表済み結果表と匿名データをつぎ合わせることによって、結果表で母集団一意に該当するレコードが誰かを特定する可能性が考えられる。そのような特定化のリスクを低減するために、公表済み結果表で度数1または2に該当する世帯が含まれるレコードを削除することによって、秘匿性を高めることができると考えられる。図5は、公表済み結果表から度数1または2に該当するレコードを削除するイメージを示したものである。

#### (6) スワッピング (data swapping)

国勢調査の匿名データを用いて作成したクロス表のセルに度数1となるレコードが発生した場合、それは、標本一意であるだけでなく母集団一意や二意に該当している可能性が考えられる。出現した母集団一意や二意が特異な属性を持つレコードである場合、個人情報特定される可能性が高くなる。また、今回は50万以上市区を提供することから、より秘匿を厳しくし、個体の特定化リスクを下げる必要がある。そこで、ある特定の地域における世帯のレコードと、他の地域の類似世帯のレコードを入れ替えることにより、個体の特定化ができないようにする。このように、非攪乱的手法だけでなく攪乱的手法も適用することによって、個体の特定が困難になり、秘匿性を高めることができる。

図6は、スワッピングのイメージを図示したものである。キー属性1とキー属性2を外観識別性の高い属性とすると、図6では、A地域とB地域において、キー属性1とキー属性2の値が一致する世帯同士を入れ替えていることがわかる。この場合、キー属性の値は同じことから、このキー属性に関するクロス表の度数は変わらないが、属性3と地域情報が入れ替わっていることから、地域と属性3に関するクロス表については、元のクロス表

図5 度数1または2に該当するレコードの削除に関するイメージ図

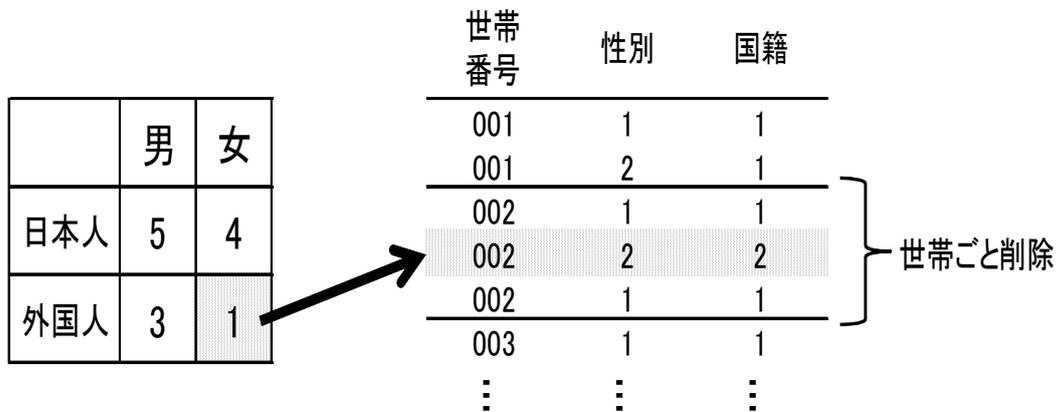
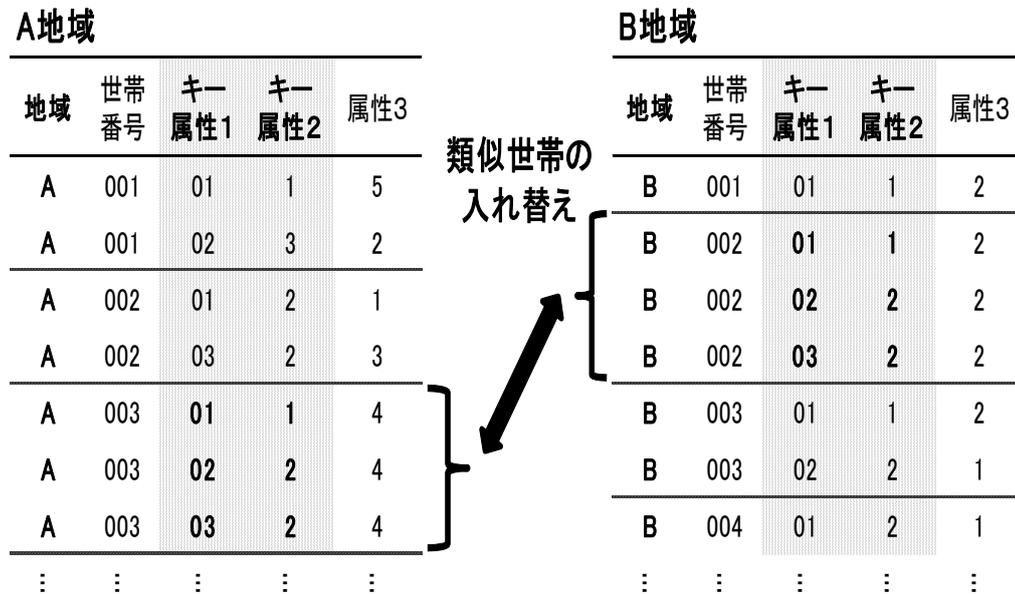


図6 スワッピングのイメージ図



と度数が異なっている。これにより個体の特定が難しくなることから、秘匿性を高めることが可能になる。

### 5. 国勢調査の匿名データの作成に関する検証

国勢調査の匿名データは、前節で述べた匿名化技法に基づいて作成されるが、その作成された匿名データの有用性と秘匿性については、定量的な評価が求められる。本節では、平成17年国勢調査の原データを用いて、使用した匿名化の方法ごとに、国勢調査の匿名デ

一タの作成に関する検証結果を述べる。

### (1) 情報の削除に関する検証

原データを用いて外観識別性の高い属性が含まれるクロス表を作成した場合、度数1または2が出現する可能性も考えられる。そのような度数1または2に該当するレコードについては、個体が特定化されるリスクが高くなる。本研究では、このようなレコードを削除するための基準を決定するために、最初に、外観識別性の高い属性を含めたクロス表を作成し、クロス表内にどの程度度数1または2が出現するかを確認する。次に、出現した度数1または2の中で特異な分布を確認した。

本研究では、世帯主と子供の年齢差に着目した。具体的には、世帯主男性と長子にあたる子供及び末子にあたる子供との年齢差のクロス表を作成した上で、年齢差の分布状況を確認した。本分析によれば、親子の年齢差が大きいまたは小さい場合、表内に度数1が多く出現する事がわかった。性別や年齢は外観識別性が高い属性であるため、個体の特定がされやすく危険である。そこで、削除の基準を決めるために、①0.5%基準に基づき、母集団の0.5%以下となるセルに該当するレコードを削除した場合 ②特異なセルの分布状況を確認し、その分布に基づいて基準を決めレコード削除をした場合を比較した。

表2は、①0.5%基準と②特異なセルの分布状況に基づく基準を適用した場合のレコード削除の閾値と削除された比率を示したものである。母集団に対する比率が0.5%以下になるセルを確認すると、長子の場合には24歳差以下と40歳差以上、末子の場合には24歳差以下と45歳差以上のセルがそれぞれ該当することが分かった。これらの該当するレコードを削除した場合、長子は13.1%、末子は4%削除されることが分かった。一方、特異なセルの分布状況を確認することによって、その分布に基づき基準を決めて削除をした場合には、長子は0.3%、末子は0.5%となり、削除対象レコード数を大幅に減少させることが可能になった。このような特異なセルの分布に基づいて削除レコード数を減らすことによって、匿名データの有用性を高めることができる。一方、特異なセルの分布状況に基づいた基準を適用することによって、リスクの高いレコードが匿名データから削除することが可能になり、秘匿性も確保されていると考えられる。

本研究では、世帯主女性と長子にあたる子供及び世帯主女性と末子にあたる子供との年齢差についても分布状況を確認した。同様に、特異なセルの分布状況に基づいた基準を適用することによって、0.5%基準を適用した場合と比べて、削除対象レコード数を相対的に減らすことが確認できた。これによって、リスクの高いレコードが匿名データから削除され、秘匿性を確保していると考えられる。

「世帯人員の多い世帯」や「世帯主又は配偶者のいずれか一方若しくは双方が外国人で子供の数が多い世帯」については、地域によって分布状況に大きな違いが見られる。これらについても、特異なセルの分布状況を確認し、地域区分ごとに削除基準を変えることで有用性を確保している。

表2 削除の基準と削除されたレコードの比率

	男親(世帯主)と長子の年齢差 削除基準と削除率		男親(世帯主)と末子の年齢差 削除基準と削除率	
	24歳差以下	40歳差以上	24歳差以下	45歳差以上
①0.5%基準による削除	9.0%	4.1%	1.9%	2.1%
②特異なセルの分布に基づき 基準を決め削除	14歳差以下	50歳差以上	19歳差以下	50歳差以上
	0.1%	0.2%	0.1%	0.4%

## (2) サンプリングに関する検証

就業構造基本調査や住宅・土地統計調査等のすでに提供されている5つの統計調査の匿名データにおいては、サンプル数が母集団の約1%になるように、住宅・土地統計調査を除く4調査については80%抽出、住宅・土地統計調査については10%抽出が行われていた。それに対して、国勢調査の匿名データの場合、原データから1%抽出であることから、誤差が大きくなる可能性が考えられる。また、世帯単位で抽出することによる人口分布の影響も考えられる。そこで、原データ、世帯単位で1%抽出したデータ、個人単位で1%抽出したデータの3データの分布を比較した(付表1)。これらの3つの分布を並べて比較したところ、大きな相違は見られなかった。したがって、世帯単位で1%抽出を行った場合、匿名データの有用性に対する影響は小さいと思われる。

## (3) 度数1または2に該当するレコード削除に関する検証

前述したとおり、セル内に度数1または2を含む公表済み結果表自体が強力な外部情報となるため、これに該当する個人が含まれる世帯のレコードは削除することが求められる。その場合、実際に母集団から削除される世帯数について把握する必要がある。

対象となる公表済み結果表から、度数1または2に該当する個人が含まれる世帯を探ることは、作業に関する時間的な制約もあることから困難であった。そこで、本研究では、外観識別性の高い7属性を選定した上で公表済みの結果表に対する擬似的なクロス表を作成し、度数1または2に該当する個人が含まれる世帯を削除した。本研究で使用した外観識別性の高い属性は、性別、世帯の種類、住居の種類、住宅の建て方、建物全体の階数、世帯が住んでいる階、床面積の7属性である。

表3-1と表3-2はそれぞれ、都道府県別と人口50万以上市区别について、世帯数に対する削除世帯数の比率を示したものである。表3-1と表3-2を見ると、度数1または2に該当する個人が含まれる世帯を削除した場合でも、削除数が最も多い地域で0.147%であることが分かった。

次に、度数1または2に該当する個人が含まれる世帯を削除したデータを用いてクロス表を作成した上で、原データのクロス表とどの程度違いが見られるかについて実験を行った。表4は、世帯類型で見た原データとレコード削除後のデータとの差の比率を示してい

表3-1 世帯数に対する削除世帯数の比率 (都道府県)

	度数1に該当するレコードの比率	度数1か2に該当するレコードの比率
北海道	0.003%	0.005%
青森県	0.009%	0.013%
岩手県	0.015%	0.024%
宮城県	0.006%	0.010%
秋田県	0.019%	0.031%
山形県	0.018%	0.029%
福島県	0.007%	0.012%
茨城県	0.007%	0.012%
栃木県	0.009%	0.015%
群馬県	0.008%	0.013%
埼玉県	0.004%	0.005%
千葉県	0.004%	0.006%
東京都	0.001%	0.001%
神奈川県	0.002%	0.003%
新潟県	0.007%	0.011%
富山県	0.009%	0.016%
石川県	0.018%	0.030%
福井県	0.018%	0.028%
山梨県	0.014%	0.024%
長野県	0.006%	0.009%
岐阜県	0.013%	0.019%
静岡県	0.005%	0.008%
愛知県	0.003%	0.005%
三重県	0.010%	0.015%
滋賀県	0.025%	0.042%
京都府	0.004%	0.006%
大阪府	0.001%	0.003%
兵庫県	0.003%	0.006%
奈良県	0.011%	0.019%
和歌山県	0.020%	0.031%
鳥取県	0.044%	0.066%
島根県	0.018%	0.031%
岡山県	0.009%	0.014%
広島県	0.008%	0.013%
山口県	0.009%	0.016%
徳島県	0.025%	0.040%
香川県	0.017%	0.029%
愛媛県	0.010%	0.015%
高知県	0.019%	0.036%
福岡県	0.003%	0.005%
佐賀県	0.030%	0.051%
長崎県	0.016%	0.025%
熊本県	0.007%	0.014%
大分県	0.010%	0.016%
宮崎県	0.016%	0.025%
鹿児島県	0.008%	0.012%
沖縄県	0.014%	0.022%

表3-2 世帯数に対する削除世帯数の比率 (50万以上市区)

	度数1に該当するレコードの比率	度数1か2に該当するレコードの比率
札幌市	0.009%	0.022%
仙台市	0.023%	0.047%
さいたま市	0.024%	0.061%
千葉市	0.030%	0.063%
船橋市	0.052%	0.112%
特別区部	0.001%	0.003%
板橋区	0.045%	0.106%
杉並区	0.024%	0.049%
大田区	0.042%	0.083%
世田谷区	0.014%	0.033%
練馬区	0.039%	0.094%
足立区	0.048%	0.098%
江戸川区	0.038%	0.080%
八王子市	0.041%	0.089%
横浜市	0.006%	0.014%
川崎市	0.021%	0.043%
相模原市	0.051%	0.094%
新潟市	0.025%	0.056%
静岡市	0.028%	0.065%
浜松市	0.030%	0.055%
名古屋市	0.009%	0.022%
京都市	0.008%	0.022%
大阪市	0.006%	0.013%
堺市	0.043%	0.094%
東大阪市	0.062%	0.147%
神戸市	0.016%	0.037%
岡山市	0.025%	0.048%
広島市	0.024%	0.054%
松山市	0.051%	0.110%
北九州市	0.025%	0.060%
福岡市	0.013%	0.027%
熊本市	0.025%	0.067%
鹿児島市	0.031%	0.062%

表4 世帯類型で見た原データとレコード削除後のデータとの差の比率

世帯類型	原データとの差(%)
夫婦のみの世帯	0.08%
夫婦と子供から成る世帯	0.07%
男親又は女親と子供から成る世帯	0.18%
三世代世帯	0.19%
その他の親族世帯	0.11%
非親族世帯	0.10%
単独世帯	0.12%

る。また、表5は、年齢と性別のクロスにおける原データとレコード削除後のデータの差の比率を表したものである。原データにおけるセルとレコード削除後のデータにおいて該当するセルの差が原データに占める比率を確認した結果、セルの差の原データに対する比率は、最も高い場合でも0.45%であることが分かった。これらの分析結果を踏まえると、度数1または2に該当する個人が含まれる世帯を削除しても、それは有用性に対して大きな影響が生じることはないと考えられる。

#### (4) スワッピングに関する検証

本研究では、有用性と秘匿性のバランスを確保する観点から、スワッピング率を変えた場合のスワッピングの有効性に関する検証を行った。スワッピングに関しては、実験に使用したキー属性や入れ替えた世帯数については、秘匿の観点から公表することはできないことから、本節では検証方法の概要のみを述べることにしたい。

本研究では、スワッピング率（入れ替える世帯の割合）を変えたテストデータ(サンプルサイズは約20万レコード)を作成した。スワッピング率は0.01%、0.1%、0.5%、1%の4パターンを設定した。原データとスワッピング済みのテストデータについて、それぞれキー属性とキー属性以外の属性のクロス表を作成するだけでなく、原データとスワッピング済みのテストデータにおける分布の差を確認することによって、有用性の比較を行った。

表6は、性別と配偶関係を例に、スワッピング率を変えた場合の原データのクロス表とスワッピング済みデータのクロス表の度数の差を示したものである。スワッピング率が上がるにつれ、度数の差が大きくなっているのが確認できる。

このように、スワッピングを適用した場合、スワッピング率の違いによって原データとスワッピング済みデータとの差を検証することは、有用性の観点から、特定のスワッピング率を決めるための重要な基礎資料となりうる。

## 6. 結びにかえて

本稿は、(独)統計センター統計技術研究課が、総務省統計局との共同研究のもとで行った国勢調査の匿名データの作成に関する実証研究の概要を述べた。諸外国の統計作成部局においては、匿名化マイクロデータの作成に関する基礎的な研究および匿名データ作成の実務への応用可能性に関する研究について多くの蓄積がある。わが国の統計作成部局においても、匿名データの作成に関する研究の蓄積が必要であると考えられており、本研究は、統計実務の観点から、統計センターで行われてきた匿名化技法に関する研究の適用可能性を追究した点で、大変意義深いと考える。また、統計センターといった統計作成部局のなかでこうした研究を行うことは、公的統計の二次利用の促進にも寄与していると思われる。

近年、公的統計のオープンデータ化、一般用マイクロデータ(仮称)の提供等、統計作成部局として公的統計のニーズに対して政策的な対応を行うことが求められているが、中長期的

表5 年齢と性別のクロスにおける原データと削除されたテストデータの差の比率

		原データとの差(%)		
		世帯主・男	世帯主・女	合計
年齢 5 歳 階 級	0~4歳	0	0	0
	5~9歳	0.00%	0.00%	0.00%
	10~14歳	0.00%	0.00%	0.00%
	15~19歳	0.00%	0.00%	0.00%
	20~24歳	0.02%	0.12%	0.06%
	25~29歳	0.06%	0.28%	0.13%
	30~34歳	0.05%	0.25%	0.10%
	35~39歳	0.10%	0.24%	0.13%
	40~44歳	0.07%	0.30%	0.11%
	45~49歳	0.05%	0.33%	0.11%
	50~54歳	0.12%	0.45%	0.18%
	55~59歳	0.06%	0.25%	0.10%
	60~64歳	0.09%	0.18%	0.11%
	65~69歳	0.04%	0.25%	0.10%
	70~74歳	0.02%	0.21%	0.08%
	75~79歳	0.08%	0.19%	0.12%
	80~84歳	0.05%	0.06%	0.05%
85歳以上	0.00%	0.20%	0.10%	

表6 スワッピング率を変えた場合の性別と配偶関係のクロス表の比較

スワッピング率0.01%

	男	女
未婚	0	-1
既婚	0	0
死別	0	0
離別	0	1
不詳	0	0

スワッピング率0.1%

	男	女
未婚	1	2
既婚	-1	-3
死別	0	-1
離別	1	5
不詳	-1	-3

スワッピング率0.5%

	男	女
未婚	-2	7
既婚	2	-3
死別	-1	2
離別	8	-2
不詳	-7	-4

スワッピング率1%

	男	女
未婚	3	5
既婚	0	-2
死別	7	26
離別	6	-23
不詳	-16	-6

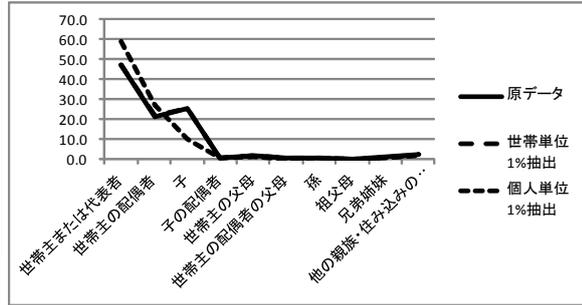
な観点から公的統計のニーズを見据えた上で研究業務を進める必要もあると考える。今後も、公的統計のマイクロデータの作成に関する基礎的な研究とともに、統計実務に対するその応用可能性に関する研究をより一層進めていきたいと考えている。

#### 参考文献

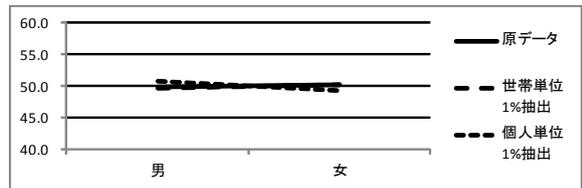
- 赤谷俊彦・荒川智浩・伊藤伸介(2014)「カナダ統計局における政府統計データの提供の動向について」、『ESTRELA』No.241, 2～9 頁
- De Kort, S., and Wathan, J.(2009) “Guide to Imputation and Perturbation in the Samples of Anonymised Records”.
- <http://www.ccsr.ac.uk/sars/resources/imputation.doc>.
- 伊藤伸介・磯部祥子・秋山裕美(2008)「匿名化技法としてのマイクロアグリゲーションの有効性に関する研究—全国消費実態調査を例に—」、『製表技術参考資料』No.10, 33～66 頁
- 伊藤伸介・磯部祥子・秋山裕美(2009)「秘匿性の評価方法に関する実証研究—全国消費実態調査のマイクロアグリゲートデータを用いて—」、『製表技術参考資料』No.11, 1～35 頁
- 伊藤伸介・高野正博・秋山裕美・後藤武彦(2010)「マイクロデータにおける有用性と秘匿性の定量的な評価に関する研究」、『製表技術参考資料』No.14, 1～40 頁
- 伊藤伸介(2011)「わが国におけるマイクロデータの新たな展開可能性について—イギリスにおける地域分析用マイクロデータを例に—」, 明海大学『経済学論集』Vol.23, No.3,36～54 頁
- 伊藤伸介・村田磨理子(2013)「家計調査マイクロデータを用いた攪乱的手法の有効性に関する研究」『製表技術参考資料』No.22, 1～26 頁
- 伊藤伸介・星野なおみ(2013)「匿名化技法としてのスワッピングの可能性について—国勢調査マイクロデータを用いた有用性と秘匿性の実証研究—」『製表技術参考資料』No.24, 1～58 頁
- 伊藤伸介(2014)「イギリスにおける政府統計データの二次的利用の現状」、『ESTRELA』No.241, 10～20 頁
- Marsh C.(1991) “The Sample of Anonymised Records” *ESRC Data Archive Bulletin*, No.48, pp.3-9.
- Marsh, C., Dale, A., Skinner, C.(1994) “Safe Data versus Safe Settings: Access to Microdata from the British Census”, *International Statistical Review*, Vol.62, No.1, pp.35-53.
- 森 博美(2000)「イギリスにおけるマイクロデータの提供」松田芳郎・濱砂敬郎・森博美編『講座マイクロ統計分析① 統計調査制度とマイクロ統計の開示』日本評論社, 48～83 頁
- Zayatz, L. (2007) “Disclosure Avoidance Practices and Research at the U.S. Census Bureau: An Update”, *Journal of Official Statistics*, Vol.23, No.2, pp.253-265.

付表1 原データと1%抽出データの分布の比較

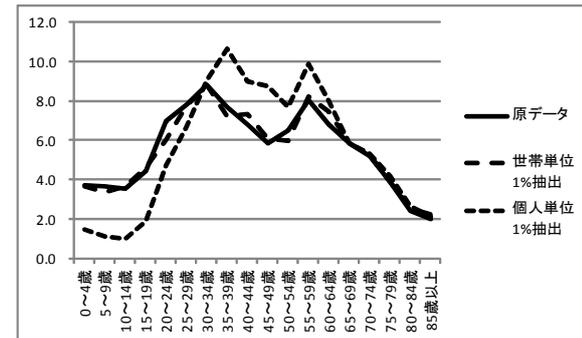
世帯主との続柄	原データ	世帯単位 1%抽出	個人単位 1%抽出
世帯主または代表者	47.3	47.2	58.8
世帯主の配偶者	21.2	21.7	27.3
子	25.5	25.1	10.1
子の配偶者	0.3	0.3	0.2
世帯主の父母	1.6	1.7	1.1
世帯主の配偶者の父母	0.4	0.4	0.1
孫	0.7	0.6	0.4
祖父母	0.0	0.0	0.0
兄弟姉妹	0.8	0.7	0.3
他の親族・住み込みの雇人・その他	2.2	2.3	1.7



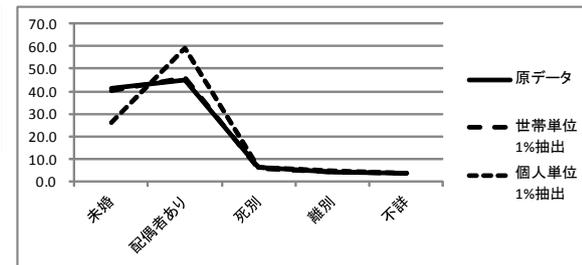
男女の別	原データ	世帯単位 1%抽出	個人単位 1%抽出
男	49.8	49.7	50.7
女	50.2	50.3	49.3



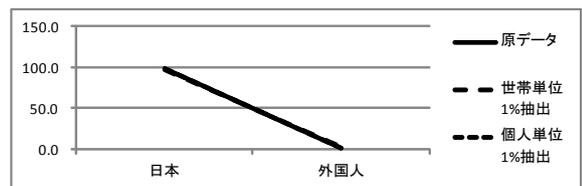
年齢5歳階級	原データ	世帯単位 1%抽出	個人単位 1%抽出
0~4歳	3.7	3.7	1.5
5~9歳	3.7	3.4	1.1
10~14歳	3.6	3.6	1.0
15~19歳	4.5	4.6	1.9
20~24歳	7.0	6.0	4.7
25~29歳	7.8	7.8	6.7
30~34歳	8.8	8.8	9.0
35~39歳	7.7	7.2	10.6
40~44歳	6.8	7.3	9.0
45~49歳	5.8	6.1	8.7
50~54歳	6.5	6.0	7.7
55~59歳	8.0	8.3	9.9
60~64歳	6.8	7.5	8.0
65~69歳	5.9	5.8	5.8
70~74歳	5.2	5.3	5.3
75~79歳	3.9	3.9	4.2
80~84歳	2.4	2.5	2.6
85歳以上	2.0	2.3	2.1



配偶者の有無	原データ	世帯単位 1%抽出	個人単位 1%抽出
未婚	41.1	40.2	26.2
配偶者あり	44.9	46.0	59.3
死別	6.0	5.9	6.1
離別	4.2	4.1	4.6
不詳	3.8	3.8	3.8

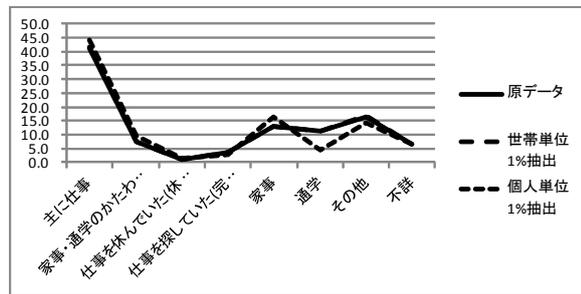


国籍	原データ	世帯単位 1%抽出	個人単位 1%抽出
日本	97.9	97.5	98.1
外国人	2.1	2.5	1.9

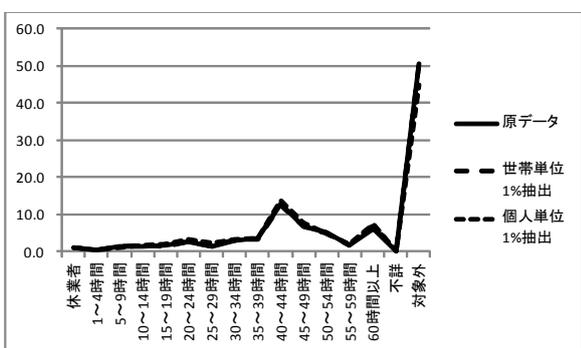


付表1 続き

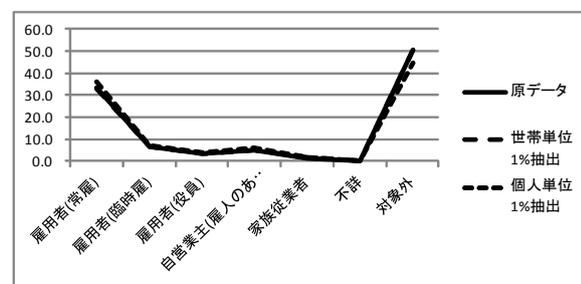
労働力状態	原データ	世帯単位 1%抽出	個人単位 1%抽出
主に仕事	41.2	41.4	44.3
家事・通学のかたわら仕事	7.3	7.3	9.7
仕事を休んでいた(休業者)	0.9	0.9	1.2
仕事を探していた(完全失業者)	3.3	3.4	2.7
家事	13.2	12.8	16.6
通学	11.2	11.2	4.4
その他	16.4	16.7	14.3
不詳	6.5	6.4	6.7



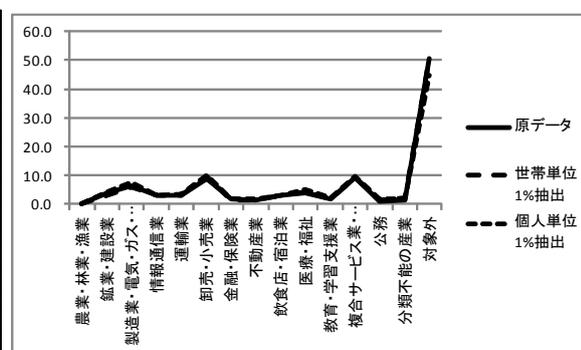
就業時間(階級)	原データ	世帯単位 1%抽出	個人単位 1%抽出
休業者	0.9	0.9	1.2
1~4時間	0.4	0.5	0.5
5~9時間	1.3	1.5	1.0
10~14時間	1.5	1.6	1.8
15~19時間	1.6	1.5	1.9
20~24時間	2.7	2.6	3.3
25~29時間	1.4	1.5	2.2
30~34時間	2.8	3.2	3.2
35~39時間	3.6	3.2	3.6
40~44時間	12.9	12.4	13.9
45~49時間	6.8	6.9	7.9
50~54時間	5.3	4.8	5.0
55~59時間	1.8	1.7	1.9
60時間以上	6.1	7.1	7.6
不詳	0.2	0.3	0.2
対象外	50.6	50.4	44.8



従業上の地位	原データ	世帯単位 1%抽出	個人単位 1%抽出
雇用者(常雇)	33.2	33.0	36.1
雇用者(臨時雇)	6.4	6.5	7.0
雇用者(役員)	3.2	3.2	4.1
自営業主(雇人のある業主)・自営業主(雇人のない業主)・家庭内職者	5.1	5.2	6.3
家族従業者	1.5	1.6	1.7
不詳	0.0	0.0	0.0
対象外	50.6	50.4	44.8

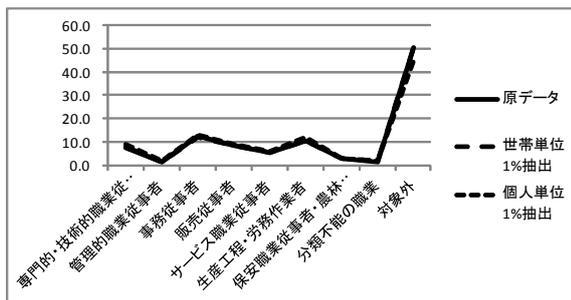


産業	原データ	世帯単位 1%抽出	個人単位 1%抽出
農業・林業・漁業	0.1	0.0	0.0
鉱業・建設業	3.8	3.3	4.1
製造業・電気・ガス・熱供給・水道業	6.4	5.9	7.5
情報通信業	3.2	3.2	3.2
運輸業	2.9	3.1	3.5
卸売・小売業	8.9	9.5	10.0
金融・保険業	1.6	1.6	1.7
不動産業	1.4	1.3	1.7
飲食店・宿泊業	3.0	3.1	3.1
医療・福祉	4.0	4.1	5.1
教育・学習支援業	1.9	1.8	2.3
複合サービス業・サービス業	9.4	9.3	9.7
公務	1.2	1.4	1.6
分類不能の産業	1.6	2.1	1.6
対象外	50.6	50.4	44.8

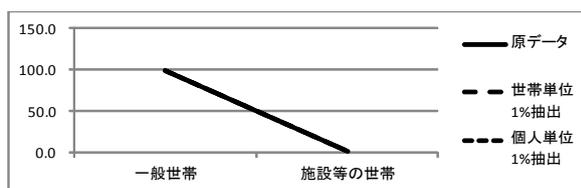


付表1 続き

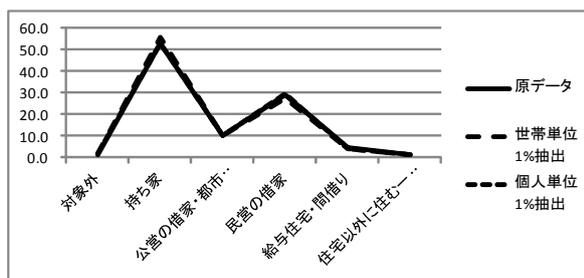
職業	原データ	世帯単位 1%抽出	個人単位 1%抽出
専門的・技術的職業従事者	7.5	7.8	8.7
管理的職業従事者	1.2	1.3	1.9
事務従事者	12.3	11.8	13.1
販売従事者	8.2	8.6	9.1
サービス職業従事者	5.4	5.2	5.9
生産工程・労務作業者	10.7	10.2	12.0
保安職業従事者・農林漁業従事者・運輸・通信従事者	2.6	2.6	2.8
分類不能の職業	1.5	2.0	1.6
対象外	50.6	50.4	44.8



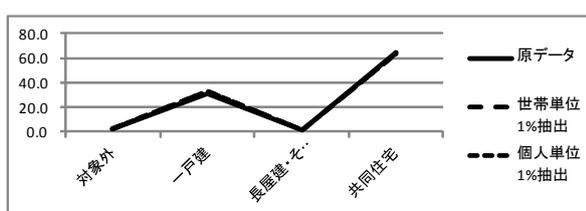
世帯の種類	原データ	世帯単位 1%抽出	個人単位 1%抽出
一般世帯	98.4	98.3	98.3
施設等の世帯	1.6	1.7	1.7



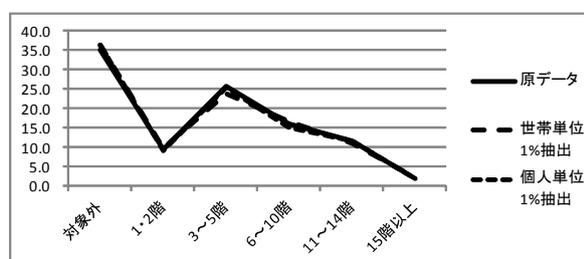
住居の種類・住宅の所有関係	原データ	世帯単位 1%抽出	個人単位 1%抽出
対象外	1.6	1.7	1.7
持ち家	53.0	55.6	53.0
公営の借家・都市機構・公社の借家	10.3	10.1	10.3
民営の借家	29.1	27.5	29.4
給与住宅・間借り	4.6	4.0	4.3
住宅以外に住む一般世帯	1.3	1.1	1.3



建て方の種類	原データ	世帯単位 1%抽出	個人単位 1%抽出
対象外	2.9	2.8	2.8
一戸建	31.3	32.6	32.6
長屋建・その他	1.0	0.9	0.9
共同住宅	64.8	63.6	63.6

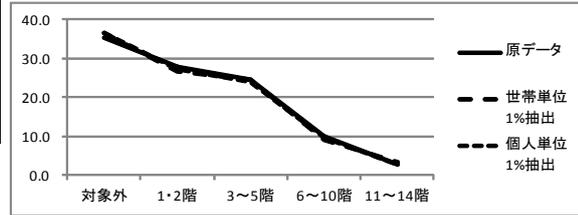


建物全体の階数	原データ	世帯単位 1%抽出	個人単位 1%抽出
対象外	35.2	36.4	36.4
1・2階	9.6	9.9	9.2
3～5階	25.8	23.9	25.4
6～10階	16.1	16.7	15.2
11～14階	11.5	11.0	11.8
15階以上	1.9	2.2	2.0

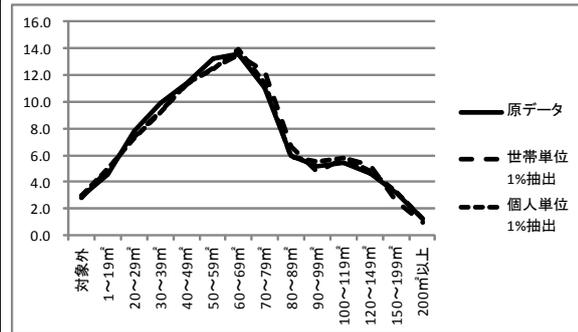


付表1 続き

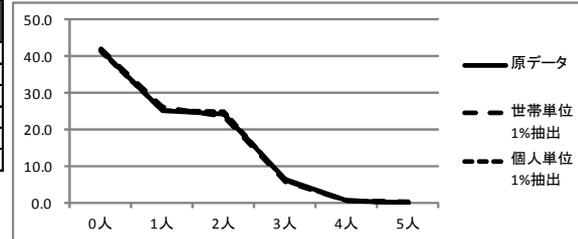
世帯が住んでいる階	原データ	世帯単位 1%抽出	個人単位 1%抽出
対象外	35.2	36.4	36.4
1・2階	27.7	26.7	27.4
3～5階	24.5	24.6	23.8
6～10階	9.8	9.1	9.6
11～14階	2.8	3.2	2.8



延べ面積	原データ	世帯単位 1%抽出	個人単位 1%抽出
対象外	2.9	2.8	3.0
1～19㎡	4.6	4.9	5.0
20～29㎡	7.8	7.3	7.3
30～39㎡	9.9	9.3	9.2
40～49㎡	11.4	11.3	11.4
50～59㎡	13.2	12.6	12.5
60～69㎡	13.6	13.4	13.9
70～79㎡	11.0	12.2	11.3
80～89㎡	5.9	6.7	5.9
90～99㎡	5.1	4.7	5.5
100～119㎡	5.5	5.8	5.8
120～149㎡	4.6	5.3	4.8
150～199㎡	3.2	2.6	3.2
200㎡以上	1.3	1.0	1.2



子供の数	原データ	世帯単位 1%抽出	個人単位 1%抽出
0人	41.6	42.0	41.3
1人	25.4	26.1	25.3
2人	24.3	23.5	24.7
3人	6.3	5.9	6.2
4人	0.8	0.6	0.8
5人	0.0	0.2	0.0



---

製 表 技 術 参 考 資 料 27

平成 26 年 10 月 発行

編 集 ・ 発 行 独 立 行 政 法 人 統 計 セ ン タ ー

〒162-8668

東京都新宿区若松町 19-1

電 話 代 表 03 ( 5273 ) 1200

---

掲載論文を引用する場合は、事前に下記まで連絡してください

統計情報・技術部統計技術研究課 TEL : 03-5273-1368

E-mail : [research@nstac.go.jp](mailto:research@nstac.go.jp)