

国勢調査の匿名化マイクロデータの作成方法に関する新たな取り組み

NSTAC

Working Paper No.36

平成 29 年 12 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. 本稿の背景と目的	3
2. スワッピング技法の有効性の評価	4
3. PRAM の可能性について	9
4. おわりに	12
参考文献	12
表1～表9	14
付表1	20
付表2	22
付表3	24
付表4	26

国勢調査の匿名化マイクロデータの作成方法に関する

新たな取り組み

伊藤伸介*、星野なおみ**、阿久津文香***、菊池亮****

要 旨

近年、「統計改革」が注目される中で、公的統計データの作成・提供への注目が高まっている。また、複数のタイプの公的統計マイクロデータの提供可能性も議論されている。

現在、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、労働力調査、および国勢調査の6つの調査の匿名データが作成・提供されている。国勢調査については、小地域の匿名データのニーズが高いことなどを踏まえ、スペシャルユニーク（特殊な一意、special uniques、ex. 全国レベルの結果表のセルで度数1になるケース）等において有益な匿名化処理の方法が模索されてきた。そうした中で、秘匿性の観点からパータベーション(攪乱的手法)のさらなる可能性を追究するだけでなく、パータベーションが適用された匿名化マイクロデータの有用性を検証することによって、ニーズに応じた匿名化マイクロデータの作成が可能になることが考えられる。そこで、本稿では、国勢調査の匿名化マイクロデータの作成方法に関する新たな取り組みとして、パータベーションの一手法であるスワッピングと PRAM (=Post Randomization Method)に焦点を当て、攪乱的手法が適用された匿名化マイクロデータの可能性を追究した。

本研究の結果を踏まえると、複数の地域間におけるスワッピングの場合、属性値がより近いレコードとの入れ替えを行うことが可能なことが明らかになった。また、年齢、産業と職業における情報量損失を小さくするために、年齢、産業と職業を層別した上で、スワッピング技法を適用することが可能なことが確認された。一方、PRAM についても年齢、産業と職業を層別した上での攪乱が可能であることが確認されたが、再構築は適用されていないことから、今後さらなる研究・実験が必要になると思われる。

* (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 (中央大学経済学部教授)

** (独) 統計センター統計情報・技術部情報ソリューション課

*** 総務省統計局国勢統計課

**** (独) 統計センター統計情報・技術部統計作成支援課非常勤研究員 (NTTセキュアプラットフォーム研究所)

国勢調査の匿名化マイクロデータの作成方法に関する

新たな取り組み

伊藤 伸介、星野 なおみ、阿久津 文香、菊池 亮

1. 本稿の背景と目的

近年、「統計改革」が注目されており、EBPM (=Evidence Based Policy Making, 客観的な事実に基づいて政策立案を行うこと) をわが国でも展開するために、公的統計のデータだけでなく、民間のビックデータや行政記録データの利用可能性が指摘されている。2017年5月に刊行された『統計改革推進会議最終取りまとめ』では、EBPMの推進体制の構築のために、「統計等データ」の整備・改善が必要であることを明記している。「統計等データ」とは、統計、統計マイクロデータ及び統計的な利活用を行うために用いられる行政記録情報であって、それらのデータの利用・解釈を行うために必要な関連情報(メタデータ)を含む。こうした「統計改革」という形で、公的統計データの作成・提供への注目が高まっている。また、いわゆる public use file を含む複数のタイプの公的統計マイクロデータの提供可能性も議論されている。

現在、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(調査票A、調査票B)、労働力調査、および国勢調査の6つの調査の匿名データが作成・提供されている。例えば、わが国の国勢調査の匿名データの特徴としては、①提供年次は、平成12年と平成17年の2回分であること、②地域区分は、都道府県と人口50万以上市区であること、③データ量は母集団の1%であり、世帯単位で抽出されていること、④1種類のみ匿名データが作成・提供されていること、⑤リコーディング、トップコーディング、レコード削除といった非攪乱的手法(non-perturbative methods)だけでなく、スワッピングといった攪乱的手法(perturbative methods、パータベーション)が適用されていることが指摘される。

国勢調査の匿名データの実務の観点から見て、小地域の匿名データのニーズが高いことなどを踏まえ、スペシャルユニーク(特殊な一意、special uniques、ex. 全国レベルの結果表のセルで度数1になるケース)等において有益な匿名化処理の方法が模索されている。匿名化マイクロデータにおける攪乱的手法の適用可能性については、さらなる検討の余地があると思われる。

これまで、匿名化マイクロデータ(匿名化技法が適用されたマイクロデータ)の作成に関する基礎研究として、国勢調査の個票データを用いて、スワッピングといった攪乱的手法の適用可

能性、さらにはリコーディングやトップコーディングにおける区分統合の可能性を検討してきた(伊藤・星野(2014), 伊藤ほか(2016a, 2016b))。これらの匿名化手法を適用することによって、様々なタイプの匿名化マイクロデータが作成される。秘匿性の観点からパータベション(攪乱的手法)のさらなる可能性を追究するだけでなく、パータベションが適用された匿名化マイクロデータの有用性を検証することによって、ニーズに応じた匿名化マイクロデータの作成が可能になる。

本稿では、国勢調査の個票データから作成された匿名化マイクロデータの展開可能性を追究する。さらに、本研究では、国勢調査のテストデータを作成した上で、攪乱的手法の適用可能性を探る。

2. スワッピング技法の有効性の評価

スワッピング(data swapping)とは、マイクロデータに含まれるレコード同士で属性値を入れ替えることであって(Willenborg and Waal(2001, p.126))、スワッピングの可能性に関する議論については少なくとも1970年代に遡ることができる(Dalenius and Reiss(1978))。スワッピングのイメージは、下図のとおりである。

原データ					匿名化マイクロデータ				
番号	地域	性別	雇用形態	週間就業時間	番号	地域	性別	雇用形態	週間就業時間
1	1	1	2	1	1	1	1	2	1
2	1	2	1	2	2	2	2	1	2
3	1	1	1	4	3	1	1	1	4
4	1	1	3	1	4	1	1	3	1
5	1	1	2	3	5	1	1	2	3
6	1	1	3	2	6	1	1	3	2
7	2	2	1	2	7	1	2	1	2
8	2	1	1	4	8	1	1	1	4
9	2	2	2	3	9	2	2	2	3

入れ替え

性別 1:男 2:女
 地域 1:三大都市圏 2:それ以外
 雇用形態 1:正規の職員・従業員 2:パート・アルバイト 3:派遣・契約社員
 週間就業時間 1:35時間未満 2:35～48時間 3:49～59時間 4:60時間以上

原データ : 秘匿処理を施していない個票データ

匿名化マイクロデータ : 原データに匿名化技法を適用することによって作成したマイクロデータ

出所 伊藤・星野(2014)

地域のみ入れ替えを行っているので、上記の例の場合、スワッピング後の匿名化マイクロデータにおいて作成された性別、雇用形態、週間就業時間別のクロス表は、スワッピング前の原データにおけるクロス表の数値と変わらない。諸外国におけるスワッピングの適用事例としては、以下の例を指摘することができる。アメリカでは、2000年人口センサスのPUMS(Public Use Microdata Samples)やAmerican Community Surveyにおいて、スワッピングを適用している。スペシャルユニークの対象となるレコードを探索した上で、別のレコードに置き換えるという処理を行っている。具体的には、非常に粗い地域区分において

特定の人口社会的属性群に基づいて一意性を有する世帯のレコードは、露見リスクが非常に高いと考えられることから、別の地域における他の世帯との入れ替えが行われている(Zayatz(2007, p.257))。また、イギリスでも、人口センサスの集計結果表を作成する前の個票データにおいて、レコードスワッピングが適用されている(Shlomo(2007))。なお、アメリカにおいても、2000年人口センサスの集計表に対して秘匿処理を行うために、人口センサスの個票データにスワッピングを適用していることが知られている。

スワッピング技法を適用する場合、2つの地域間でレコードの入れ替えを行うことが想定されるが、スワッピングにおける秘匿性および有用性を勘案した場合、 n 地域間でスワッピングを行うことも考えられる。近年では、アメリカセンサス局において、 n -サイクル スワッピング(n -cycle swapping)と呼ばれる複数地域間のスワッピングの考え方も提唱されている(Depersio et al.(2012))。そこで、本研究では、複数地域間のスワッピングの可能性を追究する。

つぎに、本研究におけるキー変数の選定方法について説明することにした。本研究では、A県を対象に、平成12年、17年のすでに提供されている国勢調査の匿名データにおける秘匿性のレベルを超えない形で、様々なリコーディング(区分統合)が適用されたテストデータを対象に、母集団一意(population unique)の比率を計測した上で、スワッピングの対象となるキー変数の選出を行った。また、地域を含む変数の原区分において、匿名データと同じ区分統合を行った場合の母集団一意の減少率も参照しながら、キー変数の区分の設定を行った。さらに、A県における地域5区分を対象に、「年齢5歳階級90歳以上トップコーディング&産業16区分&職業7区分」の中で最も母集団一意が低い組合せを選出した。なお、本研究における地域5区分とは、(1)地域A(人口は50万人以上の地域)、(2)地域B(人口は50万人以上の地域)、(3)地域C(人口は20万人以上の地域)、(4)地域D(人口は20万人以上の地域)、および(5)その他(地域A、地域B、地域C、地域D除いたA県における地域)である。なお、地域A、地域B、地域Cと地域Dの順に、人口規模は小さくなっている。

本研究において使用するキー変数および変数の分類区分はつぎのとおりである。

- ①建物の建て方(3区分)
- ②住居の種類(5区分)
- ③性別(2区分)
- ④配偶者の有無(4区分)
- ⑤国籍(2区分)
- ⑥労働力状態(7区分)
- ⑦従業上の地位(5区分)
- ⑧年齢(19区分)
- ⑨産業大分類(16区分)
- ⑩職業大分類(7区分)

本研究においては、地域に関して、人口50人以上、あるいは人口20万人以上の地域規

模に該当する、地域 A、地域 B、地域 C と地域 D の 4 地域が選ばれた。

つぎに、本研究におけるスワッピングの方法は、伊藤・星野(2014)を参考にしながら、4 地域(地域 A、地域 B、地域 C、地域 D)を対象として、以下の手順でスワッピングを行う。

- ① スワッピングを施すにあたって、年齢、産業と職業についてグループ化を行う。本研究では、異なる年齢層、産業区分や職業区分の入れ替えに伴う情報量損失を小さくすることを指向していることから、具体的には、年齢については 10 歳区分、産業については 3 区分(第 1 次産業、第 2 次産業、第 3 次産業)、職業については 2 区分(ブルーカラーとホワイトカラー)でグループ化を行った。なお、ホワイトカラーについては、専門的・技術的職業従事者、管理的職業従事者と事務従事者が該当するものとし、ブルーカラーについてはそれら以外の分類区分が該当するとする。
- ② スワッピングの対象レコードの中で優先順位が高いレコードを探索する。それは、つぎのように行われる。キー変数から選ばれた任意の 3 変数のクロス表においてスペシャルユニークに該当する回数が多いレコードを、スワッピングの対象とする。具体的には、ある特定のレコードがスペシャルユニークに該当した回数をレコードごとに計測し、スペシャルユニークの回数を点数で表す(例えば、10 個の 3 変数の組み合わせにおけるクロス表でスペシャルユニークに該当したのであれば、10 点とする等)。点数が高いほど、スペシャルユニークの中でもリスクが高いレコードと考えることができることから、スワッピングの優先順位が高くなる。なお、年齢、産業と職業の少なくともいずれかが含まれるキー変数の組み合わせの優先度が高くなっている。
- ③ ②で求めたスペシャルユニークに該当する点数に基づいて、地域ごとに、スペシャルユニークの点数が高いレコードから順番に、一定のスワッピング率にしたがってスワッピングを実行する(ターゲット・スワッピング)。本研究では、スワッピング率を 0.1%とした。また、スワッピングにおいては、地域 A、地域 B、地域 C、および地域 D の順番で行った。具体的には以下のとおりである。最初に、地域 A のスペシャルユニークを地域 B から探して入れ替える(ドナーファイルにおいて 1 対 1 で対応するレコードと入れ替えを行う)。入れ替えられたレコードをスワッピングのドナーのレコードとはしない。それを繰り返して、 $p\%$ (本研究では $p=0.1\%$)のスワッピング率に達するまでスワッピングを行う。
- ④ 年齢、産業と職業で層化されたレコード群を対象に、ドナーファイルから層内で同一の属性値の組を有するレコードを探索した上で入れ替えを行う。層内において 1 対 1 で対応するレコードが存在しない場合には、ドナーファイルにおいて同一の属性値の組を有する n 個のレコードの中からランダムに選定した上で入れ替えを行う。さらに、ドナーファイルにおいて同一の属性値の組を有するレコードがない場合には、分類区分数の逆数をウェイトとした上で、距離の計算を行い、年齢、産業と職業のグループ内で距離が一番近いレコードと入れ替える。

つぎに、本研究における距離の計算方法は、以下のとおりである(伊藤・星野(2014))。

(1)キー変数 10 変数について、スワッピングの対象レコードとドナーファイルの中のレコードの間で一致するかどうか検討する。具体的には、本研究では、それぞれスワッピングの対象レコードに含まれる属性値とドナーファイルの中のレコードにおける値が一致する場合には0、それ以外には1というスコアを新たに設定した上で、分類区分の区分数で割る(不詳については除外している)。なお、年齢、産業、職業については、ドナーの層に含まれる区分数で除する。

(2)上記の 10 変数に関する値を合計して距離を計測する。距離の計測式は、以下のよう示される。

$$\begin{aligned} \text{距離} = & \text{[建物の建て方の分類区分数の逆数]} \times \text{[建物の建て方のスコア]} \\ & + \text{[住居の種類分類区分数の逆数]} \times \text{[住居の種類スコア]} \\ & + \text{[性別スコア]} \\ & \dots \\ & + \text{[職業分類区分数の逆数]} \times \text{[職業スコア]} \end{aligned} \quad (1)$$

なお、同じ距離のレコードが複数ある場合には、ランダムに1つのレコードを選択している。

他方、本研究では、スワッピングの精度の評価に関してつぎの2つの方法を行った。

第1に、3変数の組み合わせのそれぞれにおいて探索されたスペシャルユニーク(3変数のクロス表における度数1)の中で、どの程度スワッピングが適用されているかについて検証を行った。それによって、どの変数に対してスワッピングが効果的に適用されているのかを確認した。

第2に、個票データとパータベーション(攪乱的手法)が適用されたデータ(この場合は、スワッピングが適用されたデータ)との間の絶対距離の平均値(average absolute distance)を計測することによって、情報量損失の計測を行った。

第2の情報量損失に基づく有用性の検証であるが、マイクロデータにおける有用性の定量的な評価方法については、クラメールのVといった関連性の指標の算出や秘匿処理が適用される前の個票データ(以下、「原データ」)からの絶対距離の平均値(average absolute distance)の計測等(伊藤・星野(2014))が考えられる。また、伊藤・星野・阿久津(2016a)においては、エントロピーに基づいて情報量損失の指標を作成した上で、秘匿の観点から「許容可能」な分類区分の組み合わせについて情報量損失の計測を行った。そこで、本研究においても、パータベーション(perturbation、攪乱的手法)が適用されたデータ(ex. スワッピング済データ)を行った場合の距離を計測することによって、情報量損失の計算を行った。具体的には、(2)式のように、原データとパータベーションが適用されたデータの両方で集計表を作成した上で、セルごとの度数の差の絶対値に関する平均値を計測した(Shlomo et

al.(2010))。

$$IL = \frac{\sum_c |T^P(c) - T^O(c)|}{n_T} \quad (2)$$

$T^O(c)$: 原データを用いて作成したクロス表におけるセルの度数

$T^P(c)$: パータバージョンが適用されたデータをもとに作成したクロス表におけるセルの度数

n_T : 集計表におけるセルの数

表1～表4はそれぞれ、地域A～地域Dにおけるスワッピングの結果を示したものである。地域A、地域B、地域Cと地域Dにおけるスペシャルユニークとなるレコードとスワッピングされたレコード数との関係はそれぞれ、付表1、付表2、付表3と付表4で示されている。例えば、表1を見ると、地域Aにおいて、スワッピング対象となった278レコードの中で、地域Bのドナーファイルと置き換えられたのは195レコードであって、残りの83レコードについては地域Cと地域Dから置き換えられている。また、表2を見ると、地域Bでは、スワッピング対象となった257レコードの中で、地域Cと入れ替えられたのは、129レコードのみであって、地域Dからは96レコードが入れ帰られている。さらに、地域Aのファイルですでにスワッピングのために用いられたレコードを除いたレコード群をドナーファイルとした場合、32レコードがスワッピングに用いられたことが確認できる。すなわち、複数の地域で入れ替えを行うことによって、1つの地域のみをドナーファイルとするよりも、よりスワッピングの対象レコードに近いドナーファイルのレコードと入れ替えを行うことが可能になっている。このことは、例えば、地域Aと地域Bの2地域間で入れ替えるよりも、複数の地域間で入れ替えを行ったほうがより効果的なスワッピングを行うことができることを示唆している。

つぎに、付表1～付表4は、120パターンでの3変数の組み合わせにおいて、スペシャルユニークとなっているレコードの中のどの程度がスワッピングの対象となったかを示している。それによれば、組み合わせによって、スペシャルユニークの対象となっているレコードの中でスワッピングが施されたレコードの比率が異なることがわかる。例えば、地域Aにおいて、住居の建て方、従業上の地位、産業の3変数については、スペシャルユニークである18レコードの中の16レコードにスワッピングが施されており、スワッピングされたレコードの比率は88.9%となっている。一方、年齢、産業と職業の3変数については、スペシャルユニークである198レコードの中の66レコードにスワッピングが施されており、その比率は33.3%にすぎない。これについては、年齢、産業と職業の3変数でスペシャルユニークに該当したレコードは、他の変数の組み合わせではスペシャルユニークに該当していなかった場合が少なくないことから、スワッピングの対象レコードの中で優先順位が相対的に低くなっていることが推察される。このように、スペシャルユニークに対するスワッピングの効果にはばらつきがあり、その効果は限定的であることが明らかになった。

表5は、①年齢5歳区分で90歳以上トップコーディング、産業(16区分)と職業(7区分)の3変数、②年齢10歳区分で90歳以上トップコーディング、産業(16区分)と職業(7区分)の3変数、および③年齢5歳区分で90歳以上トップコーディング、産業(3区分)と職業(2区分)でクロス表を作成した上で、地域A～地域Dを対象に、スワッピングにおける情報量損失を計測したものである。産業と職業をそれぞれ3区分、2区分と粗くした③のケースにおける情報量損失が最も低くなっていることがわかる。一方で、スワッピング率が0.1%であることから、情報量損失にそれほど大きな違いは見られないことが確認できる。

3. PRAMの可能性について

攪乱的手法には、スワッピング技法の他にも Post Randomization Method (PRAM) と呼ばれる、ノイズ付加等の確率的処理を施してプライバシーを保護する方法が知られている (Kooiman (1998))。PRAMはオランダの公的統計 (de Wolfら(1998), de Wolf and van Gelder (2004)) や、データベース (Agrawal and Srikant (2000), Agrawal et al.(2005)) の分野において幾つか適用可能性の研究がなされてきている。

PRAMは「攪乱」と「再構築」と呼ばれる2つのステップから構成される。PRAMにおける「攪乱」とは、個票データの各セルの値をあらかじめ決められた確率に基づいて遷移させるステップであり、再構築とは、攪乱された個票データから元の個票データが持つ分布を推定するステップである。

攪乱をより詳細に説明するため、幾つかの定義を導入する。個票データの各属性を V とし、その取り得る属性値を $\{v_i\}_{0 \leq i < M}$ とする。また、ある属性 V に対する遷移確率行列 A を

$$A := \begin{pmatrix} a_{0,0} & \cdots & a_{0,M-1} \\ \vdots & \ddots & \vdots \\ a_{M-1,0} & \cdots & a_{M-1,M-1} \end{pmatrix}$$

と定義する。ここで、 $a_{i,j}$ は v_i が v_j に遷移する確率であり、匿名化処理を Δ としたとき $a_{i,j} := \Pr[v_j = \Delta(v_i)]$ と書ける。

攪乱とは、各属性値を遷移確率行列 A に従って遷移させることである。例えば、性別という属性で、属性値が“男性”・“女性”の2つのみであり、遷移確率行列が

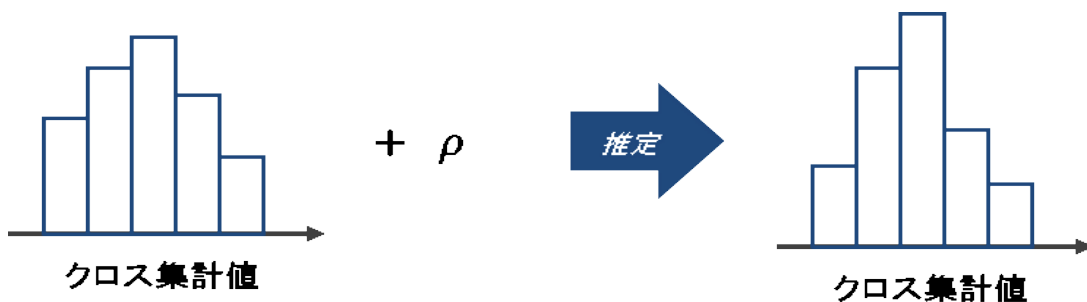
$$A := \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$$

である場合、元々属性値が“男性”であったセルは、75%の確率でそのまま維持され、25%の確率で“女性”に遷移される。

攪乱の効果は、直感的には以下のような例から得られる。攻撃者は、個票データの中から“男性かつ地域A”に住む、ある人物のデータを特定しよう試みているとする。このとき、攪乱によって“男性かつ地域A”が“女性かつ地域A”に遷移した場合、値が変わっているため攻撃者にとってその人物を特定することが難しくなっている。さらに、もし“男性かつ

地域A”が遷移せずにそのまま残ったとしても、攻撃者から見れば“男性かつ地域A”が遷移せずにそのまま残っているのか、もしくは“女性かつ地域B”という人物のデータがたまたま“男性かつ地域A”に遷移したのか、どちらなのかを判別することは難しく、結果として個票データの中からある人物を特定することは難しくなっている。このようなプライバシー保護の効果を定量化した結果として、PRAMは、高々確率 $1/k$ でしか個人を特定できないことを保証するPk-匿名性(Ikarashi et al. (2014))や、 ϵ -差分プライバシー(Dwork (2006))を満たせることが知られている(Lin et al. (2012), Ikarashi et al. (2014))。

PRAMのもう1つのステップである再構築とは、元の個票データの統計的な分布を推定し、精度を向上させる処理である。一般にPRAMでは、攪乱に用いた遷移確率行列は攪乱されたデータと共に公開される。これらの情報を用いると、元の個票データのクロス集計値などを推定することができる。例えば、先ほどの“男性”・“女性”のみの例を考えると、攪乱データの男性と女性の人数の比率は、元の個票データの男性と女性の人数の比率に比べ1:1に近づいている。そのため、例えば攪乱データで男性60人、女性40人であったならば、元の個票データでは男性70人、女性30人のように、より大きな差があったと考えられる。このような推定が再構築である。そのイメージを下図に示す。



再構築を行う具体的なアルゴリズムとしては、クロス集計値を再構築する逐次ベイズ法(Kooiman et al.(1998), Agrawal (2000))がよく知られている。また、近年では再構築によって得られた結果が元の個票データとどの程度離れているかといった精度を保証できるような手法も研究されている(長谷川(2016))。

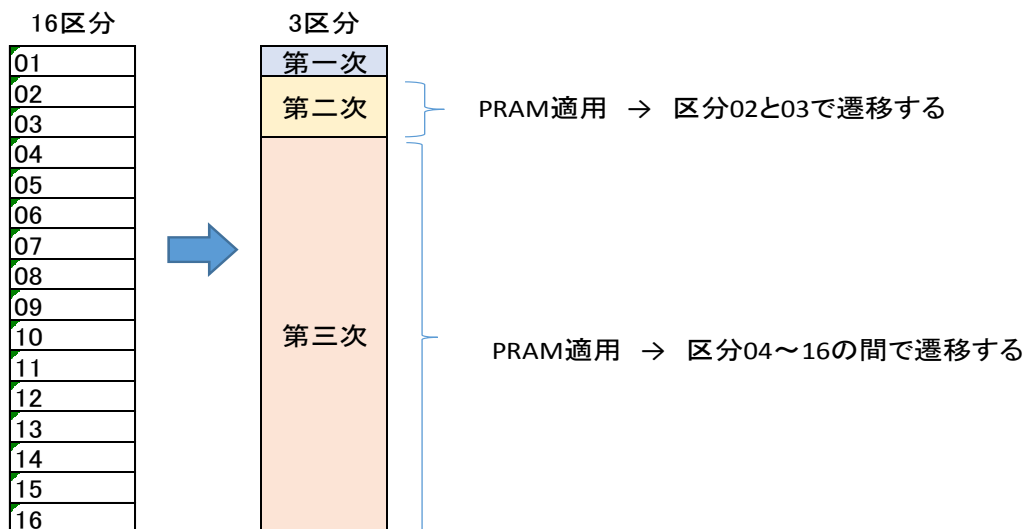
本研究では、国勢調査の個票データに基づく匿名化マイクロデータにおいて、年齢、産業と職業のそれぞれに対して攪乱を行い、その結果について考察する。

既に述べた通り、PRAMでは、ある個人の属性値 v_i が異なる属性値 v_j に変化し得ることに加えて、別の個人の属性値 v_k がたまたま v_i に変化し得ることも、プライバシーを保護する上で重要な役割を果たしている。そのためPRAMでは、ある属性値 v_i は、攪乱によってその属性の取り得る値 $\{v_i\}_{0 \leq i < M_i}$ の全てに遷移し得るのが一般的である。言い換えると、遷移確率行列の各要素について $a_{i,j} \neq 0$ である。

しかし、利用者の観点からは、そのような攪乱が適切でない場合も考えられる。例えば産業といった属性を攪乱するにあたり、第一次産業である農業と、第三次産業である卸売・小

売業といった異なる性質を持つものの中で値が遷移すると、攪乱後の分析手法によっては元の個票データを用いた場合と著しく異なる場合もある。また、そもそも第一次/第二次/第三次産業ごとに分析するといった用途では、産業区分を超えた攪乱ではなく、同一の産業区分内での攪乱に留めた方が、より分析精度が高くなることが期待できる。

そこで今回の実験における攪乱では、属性全体を攪乱する場合に加えて、属性値の幾つかをグループ化し、そのグループ内で遷移させることも行った。例えば産業の場合、16区分での攪乱に加え、下図のように、粗い3区分（第一次、第二次、第三次）にデータを分け、そのグループごとに攪乱を行った。



同様に職業は7区分と2区分の2パターン、年齢は5歳刻みと10歳刻みの2パターンで攪乱を行った。

攪乱方法には維持置換攪乱を用いた。これは、一定の確率 ρ で属性値を維持し、確率 $1-\rho$ で遷移候補の中からランダムな値に遷移するような攪乱であり、遷移確率行列の各要素は

$$a_{i,j} = \begin{cases} \rho + \frac{1-\rho}{|V|} & \text{if } i = j \\ \frac{1-\rho}{|V|} & \text{otherwise} \end{cases}$$

で与えられる。 ρ は0.95、0.9、0.85および0.8の4パターンを用いた。

本実験では地域A、地域B、地域Cと地域Dに対してPRAMを行い、精度の測定（差および差/セル数）に関しては3回の実験の平均から求めた。これらを表6～表9に示している。維持確率が大きいほど精度が高いことが確認できる。

結果を見るとスワッピングに比べ精度が低い、これはスワッピングが全レコードのうち0.1%をスワッピングしていることに対し、PRAMの実験では、維持確率が最も高い0.95であっても、全レコードのうち約15%のレコードのいずれかのセルがランダムな値に置き換わっているためと考えられる。スワッピングとPRAMの効果を正確に比較するためには、両手法の秘匿性・有用性を定量化する必要があるが、パータベーションの方法が大きく異なる

るため定量化は簡単ではなく、今後の課題である。

また、本実験では再構築を行っていない。これは、既存の再構築においては「グループに分けて攪乱する」ことを想定していないことに由来している。詳細な説明は避けるが、既存の手法をそのまま用いた場合、再構築によって産業の区分が第一次産業から第二次産業に移るといったことが起こり得るだけでなく、大量の記憶領域を必要として実行が難しいなどといったことが発生している。そのため、再構築には既存と異なる新たな手法の考案が必要となる。

4. おわりに

本稿では、国勢調査の匿名化マイクロデータの作成方法に関する新たな取り組みとして、攪乱的手法であるスワッピングと PRAM に焦点を当て、攪乱的手法が適用された匿名化マイクロデータの可能性を追究した。

本研究の結果を踏まえると、複数の地域間におけるスワッピングの場合、属性値がより近いレコードとの入れ替えを行うことが可能なことが明らかになった。また、年齢、産業と職業における情報量損失を小さくするために、年齢、産業と職業を層別した上で、スワッピング技法を適用することが可能なことが確認された。一方、PRAM についても年齢、産業と職業を層別した上での攪乱が可能であることが確認されたが、再構築は適用されておらず、今後さらなる研究・実験が必要になると思われる。

今後の研究課題として、スワッピング手法や PRAM 等のパータバージョンの統計実務における適用可能性を模索していきたい。例えば、不詳のデータに対するパータバージョンの適用については、検討課題になりうるものとする。また、パータバージョンを適用したデータの秘匿性と有用性を比較・評価するための定量的な評価方法についても、さらなる研究を進めていきたい。

参考文献

- Agrawal, R., Srikant, R. *Privacy-Preserving Data Mining*, SIGMOD 2000.
- Dalenius, T and Reiss, S. P. (1978) "Data-Swapping: A Technique for Disclosure Control (Extended Abstract)", in Proceedings of the Section on Survey Research Methods, American Statistical Association, Washington, D.C., pp.191-194.
- De Waal, T. and Willenborg, L. (1999) "Information Loss Through Global Recoding and Local Suppression", *Netherlands Official Statistics (special issue on SDC)*, Vol.14, pp.17-20.
- De Wolf, P.P. and van Gelder, I.(2004) An empirical evaluation of PRAM, Discussion paper 04012, Statistics Netherlands.
- Domingo-Ferrer, J. and Torra, V. (2001) "Disclosure Control Methods and Information

- Loss for Microdata”, Doyle *et al.*(eds.) *Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies*, Elsevier Science, Amsterdam, pp. 91-110.
- DePersio, M., Lemons, M., Ramanayake, K. A., Tsay, J., Zayatz, L.(2012) “n-Cycle Swapping for the American Community Survey”, Domingo-Ferrer, J. and Tinnirello, I.(eds) *Privacy in Statistical Databases UNESCO Chair in Data Privacy International Conference, PSD 2012 Palermo, Italy, September, 2012 Proceedings, Springer*, pp.143-164.
- Dwork, C. *Differential privacy*. ICALP, 2006.
- 長谷川聡・正木 彰伍・濱田 浩気・菊池 亮 (2016)「確率的 k-匿名化における再構築の正確度に関する理論的解析」『暗号と情報セキュリティシンポジウム』
- Ikarashi, D., Kikuchi, R., Chida, K., and Takahashi, K. “k-anonymous Microdata Release via Post Randomisation Method”, IWSEC 2014.
- 伊藤伸介・星野なおみ(2014)「国勢調査マイクロデータを用いたスワッピングの有効性の検証」『統計学』107号,1～16頁
- 伊藤伸介・星野なおみ・阿久津文香(2016a)「国勢調査における匿名化マイクロデータの有用性と秘匿性の定量的な評価」『製表技術参考資料』No.32, 1～33頁
- 伊藤伸介・星野なおみ・阿久津文香(2016b)「国勢調査マイクロデータに対する匿名化措置の可能性に関する研究」『製表技術参考資料』No.34, 1～59頁
- Kooiman, P., L. Willenborg and J. Gouweleeuw (1998) “PRAM: A Method for Disclosure Limitation of Microdata”, Research Paper, No. 9705, Statistics Netherlands, Voorburg.
- Lin, B.R., Wang, Y., and Rane, S. *A framework for privacy preserving statistical analysis on distributed databases*. WIFS, 2012.
- Shlomo, N.(2007) “Statistical Disclosure Control Methods for Census Frequency Tables”, *S3RI Methodology Working Papers* M07/04, pp.1-40.
- Shlomo, N., Tudor, C., and Groom, P. (2010) “Data swapping for protecting census tables” . J. Domingo-Ferrer and E. Magkos(eds)*Privacy in Statistical Databases UNESCO Chair in Data Privacy International Conference, PSD 2010 Corfu, Greece, September, 2010 Proceedings*, pp. 41-51. New York: Springer.
- 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.
- Willenborg, L. and de Waal, T.(2001) *Elements of Statistical Disclosure Control*, Springer, New York.

表1 地域Aにおけるスワッピングの結果

地域A	レコード数	10変数での母集団一意		スワッピング			
		レコード数	比率	対象数	うち地域Bから	うち地域Cから	うち地域Dから
	277,665	18,191	6.55%	278	195	40	43

表2 地域Bにおけるスワッピングの結果

地域B	レコード数	10変数での母集団一意		スワッピング 対象数	うち地域Cから	うち地域Dから	地域Aのスワッピング で ドナーとして入れ替え
		レコード数	比率				
	257,451	17,708	6.88%	257	129	96	32

表3 地域Cにおけるスワッピングの結果

地域C	レコード数	10変数での母集団一意		スワッピング 対象数	うち地域Dから	地域Aのスワッピング で ドナーとして入れ替え	地域Bのスワッピング で ドナーとして入れ替え
		レコード数	比率				
	85,640	8,768	10.24%	86	64	6	16

表4 地域Dにおけるスワッピングの結果

地域D	レコード数	10変数での母集団一意		スワッピング 対象数	うち地域Aから	地域Aのスワッピング で ドナーとして入れ替え	地域Bのスワッピング で ドナーとして入れ替え	地域Cのスワッピング で ドナーとして入れ替え
		レコード数	比率					
	76,442	10,011	13.10%	76	50	9	13	4

表5 スワッピングにおける情報量損失の結果

① 年齢5歳区分で90歳以上トップコーディング、産業(16区分)、職業(7区分)

	地域A	地域B	地域C	地域D
差	248	400	260	222
差/セル数	0.12	0.19	0.12	0.10

注 3変数のクロス表のセル数は2147

② 年齢10歳区分で90歳以上トップコーディング、産業(16区分)、職業(7区分)

	地域A	地域B	地域C	地域D
差	222	338	228	208
差/セル数	0.20	0.30	0.20	0.18

注 3変数のクロス表のセル数は1130

③ 年齢5歳区分で90歳以上トップコーディング、産業(3区分)、職業(2区分)

	地域A	地域B	地域C	地域D
差	10	22	12	8
差/セル数	0.08	0.17	0.09	0.06

注 3変数のクロス表のセル数は133

表6 PRAMにおける情報量損失の結果、地域A

① 年齢5歳区分で90歳以上トップコーディング、産業(16区分)、職業(7区分)

維持確率	0.95	0.9	0.85	0.8
差	1808	2985	4019	4868
差/セル数	0.84	1.39	1.87	2.27

② 年齢10歳区分で90歳以上トップコーディング、産業(16区分)、職業(7区分)

維持確率	0.95	0.9	0.85	0.8
差	1343	2453	3421	4263
差/セル数	1.19	2.17	3.03	3.77

③ 年齢5歳区分で90歳以上トップコーディング、産業(3区分)、職業(2区分)

維持確率	0.95	0.9	0.85	0.8
差	205	361	491	564
差/セル数	1.539	2.712	3.694	4.241

表7 PRAMにおける情報量損失の結果、地域B

① 年齢5歳区分で90歳以上トップコーディング、産業(16区分)、職業(7区分)

維持確率	0.95	0.9	0.85	0.8
差	1757	2811	3899	4859
差/セル数	0.82	1.31	1.82	2.26

② 年齢10歳区分で90歳以上トップコーディング、産業(16区分)、職業(7区分)

維持確率	0.95	0.9	0.85	0.8
差	1357	2297	3297	4280
差/セル数	1.20	2.03	2.92	3.79

③ 年齢5歳区分で90歳以上トップコーディング、産業(3区分)、職業(2区分)

維持確率	0.95	0.9	0.85	0.8
差	216	339	542	544
差/セル数	1.62	2.55	4.08	4.09

表8 PRAMにおける情報量損失の結果、地域C

① 年齢5歳区分で90歳以上トップコーディング、産業(16区分)、職業(7区分)

維持確率	0.95	0.9	0.85	0.8
差	1028	1717	2261	2760
差/セル数	0.48	0.80	1.05	1.29

② 年齢10歳区分で90歳以上トップコーディング、産業(16区分)、職業(7区分)

維持確率	0.95	0.9	0.85	0.8
差	783	1373	1853	2351
差/セル数	0.69	1.21	1.64	2.08

③ 年齢5歳区分で90歳以上トップコーディング、産業(3区分)、職業(2区分)

維持確率	0.95	0.9	0.85	0.8
差	153	205	250	335
差/セル数	1.15	1.54	1.88	2.52

表9 PRAMにおける情報量損失の結果、地域D

① 年齢5歳区分で90歳以上トップコーディング、産業(16区分)、職業(7区分)

維持確率	0.95	0.9	0.85	0.8
差	1115	1874	2485	3057
差/セル数	0.52	0.87	1.16	1.42

② 年齢10歳区分で90歳以上トップコーディング、産業(16区分)、職業(7区分)

維持確率	0.95	0.9	0.85	0.8
差	829	1482	2032	2599
差/セル数	0.73	1.31	1.80	2.30

③ 年齢5歳区分で90歳以上トップコーディング、産業(3区分)、職業(2区分)

維持確率	0.95	0.9	0.85	0.8
差	149	223	324	405
差/セル数	1.12	1.67	2.44	3.05

付表1 スペシャルユニークとスワッピングされたレコード数との関係、地域A

建て方 3区分	住居種 類 5区分	性別	配偶	国籍 2区分	労働力 7区分	従業上 5区分	5歳 90トッ プ 19区分	産業 16区分	職業 7区分	母集団 一意数 (a)	スワッ ピング された 数(b)	(b) / (a)
*	*	*								0	0	-
*	*		*							0	0	-
*	*			*						1	0	0
*	*				*					2	0	0
*	*					*				1	0	0
*	*						*			10	4	0.40000
*	*							*		5	1	0.20000
*	*								*	2	0	0
*		*	*							0	0	-
*		*		*						0	0	-
*		*			*					0	0	-
*		*				*				1	1	1.00000
*		*					*			5	4	0.80000
*		*						*		3	2	0.66667
*		*							*	1	1	1.00000
*			*	*						0	0	-
*			*		*					4	3	0.75000
*			*			*				3	2	0.66667
*			*				*			11	6	0.54545
*			*					*		9	7	0.77778
*			*						*	3	3	1.00000
*				*	*					2	1	0.50000
*				*		*				4	2	0.50000
*				*			*			10	6	0.60000
*				*				*		5	3	0.60000
*				*					*	3	1	0.33333
*					*	*				2	1	0.50000
*					*		*			67	41	0.61194
*					*			*		13	9	0.69231
*					*				*	6	4	0.66667
*						*	*			30	21	0.70000
*						*		*		18	16	0.88889
*						*			*	3	3	1.00000
*						*		*		18	16	0.88889
*						*			*	3	3	1.00000
*								*	*	29	20	0.68966
	*	*	*							0	0	-
	*	*		*						0	0	-
	*	*			*					0	0	-
	*	*				*				0	0	-
	*	*					*			6	5	0.83333
	*	*						*		3	2	0.66667
	*	*							*	1	1	1.00000
	*		*	*						0	0	-
	*		*		*					5	4	0.80000
	*		*			*				4	2	0.50000
	*		*				*			20	8	0.40000
	*		*					*		14	10	0.71429
	*		*						*	3	3	1.00000
	*			*	*					2	1	0.50000
	*			*		*				3	1	0.33333
	*			*			*			13	7	0.53846
	*			*				*		5	4	0.80000
	*			*					*	3	2	0.66667
	*				*	*				3	2	0.66667
	*				*		*			74	43	0.58108
	*				*			*		17	11	0.64706
	*				*				*	6	5	0.83333
	*					*	*			41	25	0.60976
	*					*		*		25	17	0.68000

付表1 続き

建て方 3区分	住居種 類 5区分	性別	配偶	国籍 2区分	労働力 7区分	従業上 5区分	5歳 90トッ プ 19区分	産業 16区分	職業 7区分	母集団 一意数 (a)	スワッ ピング された 数(b)	(b) / (a)
	*					*			*	4	2	0.50000
	*						*	*		118	51	0.43220
	*						*		*	40	28	0.70000
	*							*	*	37	28	0.75676
		*	*	*						0	0	-
		*	*		*					1	1	1.00000
		*	*			*				0	0	-
		*	*				*			5	3	0.60000
		*	*					*		0	0	-
		*	*						*	0	0	-
		*		*	*					0	0	-
		*		*		*				1	1	1.00000
		*		*			*			4	3	0.75000
		*		*				*		1	1	1.00000
		*		*					*	0	0	-
		*			*	*				0	0	-
		*			*		*			19	13	0.68421
		*			*			*		1	1	1.00000
		*			*				*	0	0	-
		*				*	*			9	7	0.77778
		*				*		*		3	3	1.00000
		*				*			*	0	0	-
		*					*	*		29	25	0.86207
		*					*		*	8	5	0.62500
		*						*	*	6	3	0.50000
			*	*	*					3	1	0.33333
			*	*		*				2	1	0.50000
			*	*			*			5	3	0.60000
			*	*				*		8	3	0.37500
			*	*					*	2	1	0.50000
			*		*	*				1	1	1.00000
			*		*		*			42	24	0.57143
			*		*			*		3	2	0.66667
			*		*				*	2	1	0.50000
			*			*	*			22	14	0.63636
			*			*		*		12	8	0.66667
			*			*			*	1	0	0
			*				*	*		65	44	0.67692
			*				*		*	29	21	0.72414
			*					*	*	21	11	0.52381
				*	*	*				0	0	-
				*	*		*			17	10	0.58824
				*	*			*		8	3	0.37500
				*	*				*	2	0	0
				*		*	*			16	9	0.56250
				*		*		*		6	5	0.83333
				*		*			*	3	1	0.33333
				*			*	*		32	20	0.62500
				*			*		*	17	11	0.64706
				*				*	*	13	7	0.53846
					*	*	*			13	10	0.76923
					*	*		*		8	6	0.75000
					*	*			*	2	1	0.50000
					*		*	*		68	36	0.52941
					*		*		*	19	16	0.84211
					*		*	*	*	31	12	0.38710
						*	*	*		110	60	0.54545
						*	*		*	34	24	0.70588
						*		*	*	36	16	0.44444
						*	*	*	*	198	66	0.33333

付表2 スペシャルユニークとスワッピングされたレコード数との関係、地域B

建て方 3区分	住居 種類 5区分	性別	配偶	国籍 2区分	労働力 7区分	従業上 5区分	5歳 90トップ 19区分	産業 16区分	職業 7区分	スペシャ ルユニ ーク数(a)	スワッ ピング された 数(b)	(b) / (a)
*	*	*								0	0	-
*	*		*							0	0	-
*	*			*						1	0	0
*	*				*					2	0	0
*	*					*				2	2	1
*	*						*			12	5	0.41667
*	*							*		9	2	0.22222
*	*								*	4	1	0.25000
*		*	*							0	0	-
*		*		*						0	0	-
*		*			*					1	1	1
*		*				*				0	0	-
*		*					*			4	4	1
*		*						*		2	2	1
*		*							*	1	1	1
*			*	*						1	0	0
*			*		*					3	1	0.33333
*			*			*				3	3	1
*			*				*			16	10	0.62500
*			*					*		14	9	0.64286
*			*						*	2	1	0.50000
*				*	*					3	2	0.66667
*				*		*				1	1	1
*				*			*			8	3	0.37500
*				*				*		7	4	0.57143
*				*					*	3	2	0.66667
*					*	*				3	2	0.66667
*					*		*			61	33	0.54098
*					*			*		9	5	0.55556
*					*				*	3	3	1
*						*	*			25	20	0.80000
*						*		*		29	26	0.89655
*						*			*	6	6	1
*						*		*		29	26	0.89655
*						*			*	6	6	1
*								*	*	36	27	0.75000
	*	*	*							0	0	-
	*	*		*						0	0	-
	*	*			*					1	1	1
	*	*				*				0	0	-
	*	*					*			8	6	0.75000
	*	*						*		2	2	1
	*	*							*	1	1	1
	*		*	*						1	0	0
	*		*		*					4	1	0.25000
	*		*			*				5	4	0.80000
	*		*				*			23	15	0.65217
	*		*					*		12	8	0.66667
	*		*						*	2	1	0.50000
	*			*	*					2	1	0.50000
	*			*		*				1	1	1
	*			*			*			11	7	0.63636
	*			*				*		11	5	0.45455
	*			*					*	5	3	0.60000
	*				*	*				4	2	0.50000
	*				*		*			87	42	0.48276
	*				*			*		10	6	0.60000
	*				*				*	2	2	1
	*					*	*			38	28	0.73684
	*					*		*		27	22	0.81481

付表2 続き

建て方 3区分	住居 種類 5区分	性別	配偶	国籍 2区分	労働力 7区分	従業上 5区分	5歳 90トップ 19区分	産業 16区分	職業 7区分	スペシ ャル ユニ ーク 数(a)	スワッ ピング され た 数(b)	(b) / (a)
	*					*			*	8	7	0.87500
	*						*	*		116	48	0.41379
	*						*		*	42	33	0.78571
	*							*	*	46	26	0.56522
		*	*	*						0	0	-
		*	*		*					0	0	-
		*	*			*				0	0	-
		*	*				*			7	6	0.85714
		*	*					*		0	0	-
		*	*						*	0	0	-
		*		*	*					0	0	-
		*		*		*				0	0	-
		*		*			*			5	5	1
		*		*				*		0	0	-
		*		*					*	0	0	-
		*			*	*				0	0	-
		*			*		*			13	11	0.84615
		*			*			*		1	1	1
		*			*				*	0	0	-
		*				*	*			9	7	0.77778
		*				*		*		3	3	1
		*				*			*	0	0	-
		*					*	*		23	19	0.82609
		*					*		*	5	4	0.80000
		*						*	*	6	5	0.83333
			*	*	*					1	1	1
			*	*		*				1	1	1
			*	*			*			9	7	0.77778
			*	*				*		8	4	0.50000
			*	*					*	3	1	0.33333
			*		*	*				0	0	-
			*		*		*			54	26	0.48148
			*		*			*		2	1	0.50000
			*		*				*	0	0	-
			*			*	*			20	14	0.70000
			*			*		*		9	7	0.77778
			*			*			*	2	1	0.50000
			*				*	*		67	36	0.53731
			*				*		*	24	20	0.83333
			*					*	*	19	11	0.57895
				*	*	*				2	0	0
				*	*		*			17	7	0.41176
				*	*			*		6	2	0.33333
				*	*				*	1	1	1
				*		*	*			8	7	0.87500
				*		*		*		6	3	0.50000
				*		*			*	3	2	0.66667
				*			*	*		40	21	0.52500
				*			*		*	9	9	1
				*				*	*	21	7	0.33333
					*	*	*			12	9	0.75000
					*	*		*		8	4	0.50000
					*	*			*	0	0	-
					*		*	*		74	32	0.43243
					*		*		*	15	10	0.66667
					*			*	*	26	13	0.50000
						*	*	*		107	57	0.53271
						*	*		*	29	17	0.58621
						*		*	*	40	19	0.47500
						*	*	*	*	187	56	0.29947

付表3 スペシャルユニークとスワッピングされたレコード数との関係、地域C

建て方 3区分	住居種類 5区分	性別	配偶	国籍 2区分	労働力 7区分	従業上 5区分	5歳 90トップ 19区分	産業 16区分	職業 7区分	スペシ ャル ユニ ーク 数(a)	スワッ ピング され た 数(b)	(b) / (a)
*	*	*								1	0	0
*	*		*							0	0	-
*	*			*						0	0	-
*	*				*					10	3	0.300
*	*					*				5	1	0.200
*	*						*			16	6	0.375
*	*							*		8	1	0.125
*	*								*	3	0	0
*		*	*							1	1	1.000
*		*		*						0	0	-
*		*			*					3	3	1.000
*		*				*				1	1	1.000
*		*					*			4	4	1.000
*		*						*		8	6	0.750
*		*							*	1	0	0
*			*	*						1	0	0
*			*		*					11	9	0.818
*			*			*				5	3	0.600
*			*				*			22	12	0.545
*			*					*		16	4	0.250
*			*						*	9	3	0.333
*				*	*					3	2	0.667
*				*		*				2	0	0
*				*			*			11	5	0.455
*				*				*		10	5	0.500
*				*					*	4	0	0
*					*	*				4	3	0.750
*					*		*			59	21	0.356
*					*			*		12	6	0.500
*					*				*	6	3	0.500
*						*	*			38	22	0.579
*						*		*		27	18	0.667
*						*			*	7	5	0.714
*						*		*		27	18	0.667
*						*			*	7	5	0.714
*								*	*	38	16	0.421
	*	*	*							1	1	1.000
	*	*		*						0	0	-
	*	*			*					2	2	1.000
	*	*				*				3	1	0.333
	*	*					*			8	6	0.750
	*	*						*		6	3	0.500
	*	*							*	1	0	0
	*		*	*						0	0	-
	*		*		*					15	8	0.533
	*		*			*				5	3	0.600
	*		*				*			27	14	0.519
	*		*					*		22	3	0.136
	*		*						*	8	3	0.375
	*			*	*					7	2	0.286
	*			*		*				0	0	-
	*			*			*			15	6	0.400
	*			*				*		10	4	0.400
	*			*					*	4	0	0
	*				*	*				2	1	0.500
	*				*		*			82	29	0.354
	*				*			*		18	5	0.278
	*				*				*	6	2	0.333
	*					*	*			47	22	0.468
	*					*		*		36	15	0.417

付表3 続き

建て方 3区分	住居種類 5区分	性別	配偶	国籍 2区分	労働力 7区分	従業上 5区分	5歳 90トップ 19区分	産業 16区分	職業 7区分	スペシャ ルユニ ク数(a)	スワッ ピング された 数(b)	(b) / (a)
	*					*			*	12	6	0.500
	*						*	*		183	55	0.301
	*						*		*	58	28	0.483
	*							*	*	59	19	0.322
		*	*	*						0	0	-
		*	*		*					1	1	1.000
		*	*			*				0	0	-
		*	*				*			8	3	0.375
		*	*					*		1	0	0
		*	*						*	0	0	-
		*		*	*					0	0	-
		*		*		*				1	0	0
		*		*			*			3	2	0.667
		*		*				*		7	6	0.857
		*		*					*	0	0	-
		*			*	*				0	0	-
		*			*		*			13	4	0.308
		*			*			*		6	1	0.167
		*			*				*	0	0	-
		*				*	*			12	10	0.833
		*				*		*		7	3	0.429
		*				*			*	1	0	0
		*					*	*		43	22	0.512
		*					*		*	12	9	0.750
		*						*	*	19	8	0.421
			*	*	*					2	1	0.500
			*	*		*				0	0	-
			*	*			*			12	2	0.167
			*	*				*		12	6	0.500
			*	*					*	2	0	0
			*		*	*				2	2	1.000
			*		*		*			56	13	0.232
			*		*			*		12	3	0.250
			*		*				*	3	1	0.333
			*			*	*			35	15	0.429
			*			*		*		20	7	0.350
			*			*			*	3	2	0.667
			*				*	*		95	28	0.295
			*				*		*	30	10	0.333
			*					*	*	43	5	0.116
				*	*	*				0	0	-
				*	*		*			19	6	0.316
				*	*			*		9	3	0.333
				*	*				*	4	0	0
				*		*	*			13	6	0.462
				*		*		*		13	7	0.538
				*		*			*	2	0	0
				*			*	*		52	22	0.423
				*			*		*	17	8	0.471
				*				*	*	17	10	0.588
					*	*	*			17	8	0.471
					*	*		*		18	6	0.333
					*	*			*	5	1	0.200
					*		*	*		124	34	0.274
					*		*		*	34	18	0.529
					*			*	*	45	7	0.156
						*	*	*		155	37	0.239
						*	*		*	57	21	0.368
						*		*	*	51	10	0.196
							*	*	*	231	38	0.165

付表4 スペシャルユニークとスワッピングされたレコード数との関係、地域D

建て方 3区分	住居種類 5区分	性別	配偶	国籍 2区分	労働力 7区分	従業上 5区分	5歳 90トップ 19区分	産業 16区分	職業 7区分	スペシ アルユニ ーク 数(a)	スワッ ピング された 数(b)	(b) / (a)
*	*	*								0	0	-
*	*		*							1	0	0
*	*			*						1	0	0
*	*				*					6	2	0.333
*	*					*				8	1	0.125
*	*						*			24	7	0.292
*	*							*		9	0	0
*	*								*	3	0	0
*		*	*							0	0	-
*		*		*						0	0	-
*		*			*					1	1	1.000
*		*				*				1	1	1.000
*		*					*			4	3	0.750
*		*						*		7	4	0.571
*		*							*	3	1	0.333
*			*	*						0	0	-
*			*		*					12	5	0.417
*			*			*				3	3	1.000
*			*				*			17	10	0.588
*			*					*		14	2	0.143
*			*						*	3	1	0.333
*				*	*					4	1	0.250
*				*		*				2	0	0
*				*			*			10	6	0.600
*				*				*		8	4	0.500
*				*					*	1	1	1.000
*					*	*				5	2	0.400
*					*		*			61	30	0.492
*					*			*		16	7	0.438
*					*				*	9	3	0.333
*						*	*			36	22	0.611
*						*		*		28	16	0.571
*						*			*	10	7	0.700
*						*		*		28	16	0.571
*						*			*	10	7	0.700
*								*	*	41	9	0.220
	*	*	*							0	0	-
	*	*		*						0	0	-
	*	*			*					1	1	1.000
	*	*				*				2	2	1.000
	*	*					*			9	6	0.667
	*	*						*		8	4	0.500
	*	*							*	2	1	0.500
	*		*	*						0	0	-
	*		*		*					13	5	0.385
	*		*			*				6	3	0.500
	*		*				*			28	12	0.429
	*		*					*		16	3	0.188
	*		*						*	2	1	0.500
	*			*	*					9	1	0.111
	*			*		*				3	0	0
	*			*			*			10	7	0.700
	*			*				*		18	4	0.222
	*			*					*	7	1	0.143
	*				*	*				4	3	0.750
	*				*		*			83	34	0.410
	*				*			*		17	7	0.412
	*				*				*	9	3	0.333
	*					*	*			40	22	0.550
	*					*		*		36	18	0.500

付表4 続き

建て方 3区分	住居種類 5区分	性別	配偶	国籍 2区分	労働力 7区分	従業上 5区分	5歳 90トップ 19区分	産業 16区分	職業 7区分	スペシヤ ルユニーク 数(a)	スワッ ピング された 数(b)	(b) / (a)
	*					*			*	14	6	0.429
	*						*	*		159	49	0.308
	*						*		*	47	22	0.468
	*							*	*	45	9	0.200
		*	*	*						0	0	-
		*	*		*					1	1	1.000
		*	*			*				0	0	-
		*	*				*			5	4	0.800
		*	*					*		0	0	-
		*	*						*	0	0	-
		*		*	*					1	0	0
		*		*		*				0	0	-
		*		*			*			2	2	1.000
		*		*				*		3	2	0.667
		*		*					*	0	0	-
		*			*	*				0	0	-
		*			*		*			16	9	0.563
		*			*			*		2	1	0.500
		*			*				*	0	0	-
		*				*	*			11	3	0.273
		*				*		*		5	3	0.600
		*				*			*	1	0	0
		*					*	*		40	27	0.675
		*					*		*	12	11	0.917
		*						*	*	8	3	0.375
			*	*	*					4	1	0.250
			*	*		*				1	0	0
			*	*			*			9	4	0.444
			*	*				*		15	3	0.200
			*	*					*	2	0	0
			*		*	*				1	1	1.000
			*		*		*			57	17	0.298
			*		*			*		9	2	0.222
			*		*				*	1	1	1.000
			*			*	*			28	11	0.393
			*			*		*		14	9	0.643
			*			*			*	3	1	0.333
			*				*	*		94	29	0.309
			*				*		*	34	17	0.500
			*					*	*	52	10	0.192
				*	*	*				2	1	0.500
				*	*		*			31	10	0.323
				*	*			*		8	4	0.500
				*	*				*	4	1	0.250
				*		*	*			17	6	0.353
				*		*		*		15	7	0.467
				*		*			*	4	0	0
				*			*	*		60	22	0.367
				*			*		*	21	6	0.286
				*				*	*	18	8	0.444
					*	*	*			19	9	0.474
					*	*		*		19	6	0.316
					*	*			*	2	0	0
					*		*	*		119	29	0.244
					*		*		*	41	19	0.463
					*			*	*	40	6	0.150
						*	*	*		124	35	0.282
						*	*		*	53	19	0.358
						*		*	*	48	8	0.167
						*	*	*	*	216	42	0.194

製 表 技 術 参 考 資 料 36

平成 29 年 12 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