

匿名データの作成・提供に係るガイドラインの改正の方向性 (匿名データの作成関係)

第 4 匿名データの作成・提供に関する計画の公表

- 諮問時の匿名データ作成計画の明示

基幹統計調査の実施に関する統計委員会への諮問時に、併せて当該年次の匿名データの提供予定時期等を明らかにすること（別紙 1 参照）を追加

第 5 匿名データの作成

- 匿名化处理の方法

これまでの統計委員会における審議結果をベースに別途策定する「匿名データの作成に係る匿名化处理基準」（別紙 2 参照）に基づいて匿名化处理を行うことを追加

第 6 匿名データの匿名化处理の実施手順

- 統計研究研修所における検証

匿名データの作成に係る統計委員会への諮問前に、匿名化处理の妥当性等に関して、事前に統計研究研修所において検証を行うことを追加

また、当該検証のために統計研究研修所に提出する資料の内容（統計調査の基本情報、匿名データの作成方針（別紙 3 参照）、審査表（別紙 4 参照）、度数分布表等）を追加

- 統計委員会への諮問

統計委員会への諮問に当たっては、審議の重点化・効率化に資するため、統計研究研修所における検証結果等を報告することを追加

※ 統計委員会における計画的かつ効率的な審議の在り方については、別途、統計委員会において検討予定。

匿名データの作成計画【雛形】（案）

〇〇〇〇調査（平成〇〇年）について、以下のとおり匿名データの作成を行う予定である。

1 〇〇〇〇調査の匿名データを作成する理由

本調査は、《調査目的を記載》を明らかにすることを目的に世帯及びその世帯員を対象に《調査周期を記載（毎月、毎年、5年ごと、等）》実施している統計調査であり、学術研究や高等教育において利用ニーズが高い調査であることから、匿名データを作成するものである。

2 作成する匿名データの種類

（調査票が1種類の場合）

本調査の調査票の匿名データを作成する。

（調査票が2種類以上の場合）

本調査は、〇〇調査票と△△調査票の2種類で行われているが、今回は〇〇調査票の匿名データを作成する。

本調査は、〇〇調査票と△△調査票の2種類で行われているが、今回は〇〇調査票と△△調査票を結合し、匿名データを作成する。

3 匿名データの作成方法の概要

（初めて匿名データの作成を行う統計調査の場合）

匿名化処理基準を参考に、以下の技法等を組み合わせて匿名化処理を行う予定である。

- ・ 元の統計調査のレコード全てを匿名データに用いるのではなく、それに間引きを施したものをを用いる（リサンプリング）
- ・ 識別情報は、レコードから全面的に削除する。また、レコードの配列順が意味をなさないように、無作為に並べ替えを行う（識別情報の削除等）
- ・ 出現頻度が低い又は特徴的な値があるレコードは、削除する（裾切りによるレコード削除）
- ・ 極端に大きな値（又は小さな値）は、上限値（又は下限値）を設けて頭打ちにする（トップコーディング（又はボトムコーディング））
- ・ 分類事項の粒度は、詳細なものではなく、粗いものとする（グルーピング）
- ・ 攪乱手法

（匿名データの年次追加を行う統計調査の場合）

匿名化処理基準に準拠したリサンプリング、グルーピング、トップコーディング、ボトムコーディングや識別情報の削除等の匿名化処理を行う予定である。

4 提供予定時期

平成〇〇年〇月

匿名データの作成に係る匿名化処理基準について

【匿名データの作成・提供に係るガイドラインの改正イメージ】

「第5 匿名データの作成」における「2(2) 匿名化の基準」に以下の内容を記載

- 調査票情報は、統計調査の目的や規模等によってその特性が異なり、一律に匿名化の基準を設定することは困難との認識の下、これまで提供機関は、匿名化する統計調査の特性を勘案し、国立大学法人一橋大学における匿名標本データの試行的提供や諸外国の統計機関の状況等を参考に匿名化の基準を設定
- 一方、基幹統計調査については、統計法（平成19年法律第53号）第35条第2項の規定に基づき、統計委員会において、データの有用性及び匿名性の確保の観点から、匿名化の基準の妥当性を確認
- このため、既に匿名データの提供を行っている基幹統計調査については、これまでの統計委員会における審議結果をベースに、「匿名データの作成に係る匿名化処理基準」を別途定め、匿名化の基準として活用
- また、新たに匿名データの作成・提供を行う統計調査については、上記匿名化処理基準を最大限活用し、効率的かつ適切な匿名化の基準を設定

【匿名データの作成に係る匿名化処理基準（案）】

これまでの統計委員会における審議結果をベースに別添のとおり整理

匿名データの作成に係る匿名化処理基準（案）

資料4別紙2（別添）

	総務省						厚生労働省		
	国勢調査	住宅・土地統計調査	就業構造基本調査	社会生活基礎調査		全国消費実態調査	労働力調査	国民生活基礎調査	
				調査票A	調査票B			匿名データA (世帯票、健康票)	匿名データB (世帯票、健康票、所得票、貯蓄票)
リサンプリング	全世帯を母集団として1%を抽出することとし、世帯の種類（「一般世帯」及び「施設等の世帯」）ごとに、市区町村及び世帯人員等で並べ替えた上で、「一般世帯」については世帯単位で、また「施設等の世帯」については個人単位で抽出処理を行い、これらを統合	リサンプリング率10%を目安として、住宅（住宅以外の建物を含む。）単位のレコードに付与された集計用乗率の大きさに基づく確率比例抽出により抽出	世帯単位に無作為抽出率は80%	世帯単位に80%を目安に抽出	世帯を単位としてまとめた上で、単純無作為抽出で世帯を抽出 抽出率は約80%	世帯単位に80%を目安に抽出	世帯を単位に層化等確率抽出法（地域11ブロック×組符号8区分）抽出率は約80%（沖縄県については、約20%）	【第1段】国勢調査区抽出 都道府県・指定都市の国勢調査区数に比例するよう、都道府県・指定都市別に国勢調査区をリサンプリング 【第2段】世帯抽出 第1段でリサンプリングされた国勢調査区から世帯をリサンプリング 抽出率：全体の約2割	
しきい値	-	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	1%	
世帯・個人を特定できる外部情報	既存の統計表により母集団一意又は二意であることが判明しているレコードの削除 ○既存の統計表により、全国において母集団一意又は二意であることが判明しているレコードが含まれる世帯の削除 ○既存の統計表のうち、特に外観識別性が高いと考えられる以下の項目が含まれる統計表について、地域（都道府県、人口50万人以上の市区）において母集団一意又は二意であることが判明しているレコードが含まれる世帯の削除	なし	なし	なし	なし	なし	なし	なし	
データの並べ替え	抽出した世帯を世帯単位でランダムに並べ替え	同一世帯の世帯員は世帯員番号順を保ったまま世帯順はランダムに並び替え	同一世帯の世帯員は世帯員番号順を保ったまま世帯順はランダムに並び替え	世帯単位にランダムに並び替え	世帯を単位としてまとめた上で乱数により並び替え	世帯をランダムに並び替え	世帯単位にランダムに並び替え	世帯順は世帯単位に乱数によりランダムに並び替え、その後データの世帯番号を付与	
世帯・個人識別情報の匿名化	地域区分ごとの出現頻度により、世帯人員が7人以上～9人以上いる世帯（世帯人員が多い世帯）を削除 未婚、死別又は離別の男親と、その未婚の20歳未満の子供のみからなる世帯（父子世帯）を削除 年齢差が25歳以上の夫婦のいる世帯（年齢差の大きい夫婦のいる世帯）を削除 年齢差が55歳以上の男親と子、年齢差が45歳以上の女親と子、年齢差が14歳以下の親と長子又は年齢差が19歳以下の親と末子のいる世帯（年齢差の大きい又は小さい親子のいる世帯）を削除 地域区分ごとの出現頻度により、子供の数が3人以上～7人以上の外国人世帯（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世帯）を削除	世帯人員が8人以上いる世帯を削除 同一年齢の15歳未満の世帯人員が3人以上いる世帯を削除 家計を支える者の年齢が15歳未満の世帯を削除	世帯人員が8人以上いる世帯を削除 同一年齢の15歳未満の世帯人員が3人以上いる世帯を削除	世帯人員が8人以上いる世帯を削除 同一年齢の15歳未満の世帯人員が3人以上いる世帯を削除	世帯人員が8人以上いる世帯を削除 同一年齢の15歳未満の世帯人員が3人以上いる世帯を削除 子どもの数の多い世帯において、地域区分ごとに発生頻度の少ない世帯 母子世帯、父子世帯において、地域区分ごとに発生頻度の少ない世帯を削除	世帯人員が8人以上いる世帯を削除 同一年齢の15歳未満の世帯人員が3人以上いる世帯を削除	世帯人員が8人以上いる世帯を削除 同一年齢階級の15歳未満の世帯人員が3人以上いる世帯を削除 転出及び死亡のレコードを削除 自衛官及び受刑者のレコードを削除	世帯人員が8人以上いる世帯を削除 同一年齢階級の世帯人員が4人以上いる世帯を削除 年齢差の大きな夫婦のいる世帯（夫が妻より20歳以上年上又は妻が夫より15歳以上年上）を削除 年齢差の大きい又は小さい親子のいる世帯（父親と末子の年齢差が55歳以上、母親と末子の年齢差が50歳以上、父親と長子の年齢差が15歳以下、母親と長子の年齢差が10歳以下）を削除 手助けや見守りを要する者が2人以上いる世帯を削除 要介護認定を受けている者が2人以上いる世帯を削除	
攪乱処理	一部世帯を同一都道府県内の他の地域の類似世帯と入れ替えるスワッピングを実施	なし	なし	なし	なし	なし	なし	なし	なし
集計用乗率（母集団推計用ウエイト）	なし	抽出率に応じて再付与して提供	提供	提供	提供	再付与して提供	提供（沖縄県については、4倍の集計用乗率を再付与）	全国一律の拡大乗数を再付与して提供 提供しない（推計世帯数の算出を前提としていないため）	
地域情報	都道府県市区（人口50万人以上のみ） 人口50万人未満の市区町村コード、調査区番号、世帯番号、調査区内連番などを削除	都道府県	3大都市圏か否か	3大都市圏か否か	3大都市圏か否か	3大都市圏か否か	提供しない	提供しない	
		市区町村番号、3大都市圏、調査単位区番号を削除	都道府県・市区町村番号、調査区番号、世帯番号を削除	調査区番号（都道府県番号、県内一連番号）世帯番号を削除	調査区番号（都道府県番号、県内一連番号）世帯番号を削除	都道府県市区町村番号、調査単位区番号、世帯一連番号を削除	調査区番号（都道府県番号、県内一連番号）、標本符号（層符号、組符号、地域符号）、世帯符号（抽出単位番号、単位内世帯番号）を削除	都道府県、地区番号、単位区番号、世帯番号を削除	

匿名データの作成に係る匿名化処理基準（案）

資料4別紙2（別添）

	総務省					厚生労働省			
	国勢調査	住宅・土地統計調査	就業構造基本調査	社会生活基礎調査		全国消費実態調査	労働力調査	国民生活基礎調査	
				調査票A	調査票B			匿名データA (世帯票、健康票)	匿名データB (世帯票、健康票、所得票、貯蓄票)
提供項目等 (その1)	世帯人員： 施設等の世帯の世帯人員は提供しない						15歳未満世帯人員： 15歳未満の男女別総数を合算し、15歳未満の総数として提供、男女別総数は提供しない		
		住宅の居住室数： 都道府県別に出現頻度により、10室以上～17室以上のいずれかでトップコーディング 世帯ごとの居住室数： 都道府県別に出現頻度により、10室以上～17室以上のいずれかでトップコーディング						居住室数： 10室以上をトップコーディング	
	住宅の床面積： 実数ではなく、公表統計に合わせた階級で提供することとし、「200～249㎡」と「250㎡以上」をグルーピング	建物の敷地面積（共同住宅）： 実数値ではなく、階級値で提供。 都道府県別に出現頻度により、75㎡未満～150㎡未満のいずれかでボトムコーディング、3000㎡以上をトップコーディング			住宅の床面積： 200㎡以上をトップコーディング 二人以上世帯は30㎡未満をボトムコーディング		住宅の床面積： 20㎡未満をボトムコーディング 300㎡以上をトップコーディング		
		建物の建築面積（共同住宅）： 実数値ではなく、階級値15区分で提供。都道府県別に出現頻度により、50㎡未満又は75㎡未満のいずれかでボトムコーディング、1000㎡以上～3000㎡以上のいずれかでトップコーディング 住宅の延べ面積： 都道府県別に出現頻度により、20㎡未満をボトムコーディング、250㎡以上～500㎡以上のいずれかでトップコーディング 住宅の敷地面積（一戸建・長屋建）： 実数値ではなく、階級値11区分で提供。 都道府県別に出現頻度により、25㎡未満及び50㎡未満のいずれかでボトムコーディング、700㎡以上～1500㎡以上のいずれかでトップコーディング 住宅の建築面積＝1Fの床面積（一戸建・長屋建）： 都道府県別に出現頻度により、20㎡未満又は30㎡未満のいずれかでボトムコーディング、150㎡以上又は200㎡以上のいずれかでトップコーディング			住宅の床面積のうち業務用： 二人以上世帯は150㎡以上をトップコーディング 単身世帯は100㎡以上をトップコーディング 住宅の敷地面積： 1000㎡以上をトップコーディング				
	建物全体の階数： 実数ではなく公表統計に合わせた階級（「1・2階建」、「3～5階建」、「6～10階建」、「11～14階建」及び「15階建以上」）で提供し、地域区分ごとの出現頻度により「6～10階建」以上又は「11～14階建」以上のいずれかでトップコーディング	建物の階数： 1～2階、6～7階、8～10階、11～14階及び15～19階でグルーピング 都道府県別の出現頻度により、「一戸建・長屋建」は、2階以上～4階以上のいずれかでトップコーディング 「共同住宅・その他」は、2階以下でボトムコーディング、11階以上～20階以上のいずれかでトップコーディング							
	世帯の住んでいる階： 実数ではなく公表統計に合わせた階級（「1・2階」、「3～5階」、「6～10階」、「11～14階」及び「15階建以上」）で提供し、地域区分ごとの出現頻度により「3～5階」以上、「6～10階」以上又は「11～14階」以上のいずれかでトップコーディング								

匿名データの作成に係る匿名化処理基準（案）

資料4別紙2（別添）

	総務省						厚生労働省		
	国勢調査	住宅・土地統計調査	就業構造基本調査	社会生活基礎調査		全国消費実態調査	労働力調査	国民生活基礎調査	
				調査票A	調査票B			匿名データA (世帯票、健康票)	匿名データB (世帯票、健康票、所得票、貯蓄票)
提供項目等 (その2)	年齢： 0～84歳を5歳階級でグルーピング 85歳以上をトップコーディング	年齢： 15歳未満は各歳、 15～89歳を5歳階級でグルーピング、 90歳以上でトップコーディング（平成20年までは15～84歳を5歳階級でグルーピング、85歳以上をトップコーディング）	年齢： 15歳未満は各歳 15～84歳を5歳階級でグルーピング 85歳以上をトップコーディング	年齢： 10歳未満は各歳 10～84歳を5歳階級でグルーピング 85歳以上をトップコーディング	年齢： 10歳未満は各歳 10～84歳を5歳階級でグルーピング 85歳以上をトップコーディング	年齢： 15歳未満は各歳 0～84歳を5歳階級でグルーピング 85歳以上をトップコーディング	年齢： 15歳未満を学齢によるグルーピング 15～84歳を5歳階級でグルーピング 85歳以上をトップコーディング	年齢： 15～89歳を5歳階級でグルーピング 90歳以上をトップコーディング 15歳未満を学齢（0～5歳、6～11歳、12～14歳）によりグルーピング	
			末子の年齢： 1～11歳を 1～2歳、 3～5歳、 6～8歳、 9～11歳で グルーピング 12歳以上をトップコーディング	末子の年齢： 1～11歳を 1～2歳、 3～5歳、 6～8歳、 9～11歳で グルーピング 12歳以上をトップコーディング					
	就業時間： 90時間以上をトップコーディング						月末一週間(ただし12月は20～26日)に仕事をした時間： 90時間以上をトップコーディング	1週間に仕事をした時間： 80時間以上をトップコーディング	
	産業： 「農業」、「林業」及び「漁業」をグルーピング 「鉱業」及び「建設業」をグルーピング 「製造業」及び「電気・ガス・熱供給・水道業」をグルーピング		産業： 詳細区分をグルーピング				事業の種類（産業）： 報告書の表章区分に合わせてグルーピング		
	職業： 「保安職業従事者」、「農林漁業作業者」及び「運輸・通信従事者」をグルーピング		職業分類： 詳細区分をグルーピング				本人の仕事の種類（職業）： 報告書の表章区分に合わせてグルーピング		
			就業開始時期： 現職が初職で70年以上前に現職に就いた者は70年でトップコーディング 前職の継続就業期間： 前職からの離職期間と前職の継続就業期間の合計が70年以上となる者は70年をトップコーディング 初職に就いた時期： 70年以上前に初職に就いた者は70年でトップコーディング					就業期間： 50年以上をトップコーディング	
	5年前の住居の所在地： 「他県から」及び「国外から」をグルーピング	従前の居住地： 都道府県別で提供	転居前の居住地： 都道府県名は提供しない						
	労働力状態： 平成12年は「家事のほか仕事」と「通学のかたわら仕事」をグルーピング 大規模調査である平成17年は就業者の内訳を提供しない							仕事の有無： 「主に通学で仕事あり」及び「その他」を「その他（仕事あり）」にグルーピング	
	家計の収入の種類： 「資金・給与が主な世帯」のうち、「農業収入もある世帯」及び「その他」をグルーピング 「農業収入が主な世帯」及び「農業収入以外の事業収入が主な世帯」をグルーピングし、その内訳は提供しない 「内職収入が主な世帯」及び「その他の収入が主な世帯」をグルーピングし、その内訳は提供しない							最多所得者か否か： 「最多所得者」及び「家計補助者又は被扶養者」の2区分にリコーディング	

匿名データの作成に係る匿名化処理基準（案）

資料4別紙2（別添）

	総務省					厚生労働省			
	国勢調査	住宅・土地統計調査	就業構造基本調査	社会生活基礎調査		全国消費実態調査	労働力調査	国民生活基礎調査	
				調査票A	調査票B			匿名データA (世帯票、健康票)	匿名データB (世帯票、健康票、所得票、貯蓄票)
提供項目等 (その3)						年間収入： 総額以外は提供しない 二人以上世帯は、2500万円以上をトップコーディング 単身世帯は、1000万円以上をトップコーディング		総所得： 単独世帯は1100万円以上をトップコーディング 2人以上世帯は2200万円以上をトップコーディング	
							総所得（雇用者所得）： 単独世帯は800万円以上をトップコーディング 2人以上世帯は1700万円以上をトップコーディング	総所得（公的年金・恩給）： 単独世帯は300万円以上をトップコーディング 2人世帯以上は500万円以上をトップコーディング	
							課税等の状況（税金＋社会保険）： 単独世帯は250万円以上をトップコーディング 2人以上世帯は490万円以上をトップコーディング	掛金： 単独世帯は40万円以上をトップコーディング 2人以上世帯は80万円以上をトップコーディング	
						貯蓄現在高： 総額以外は提供しない 二人以上世帯は、9500万円以上をトップコーディング 単身世帯は、5500万円以上をトップコーディング	貯蓄現在高： 単独世帯は6300万円以上をトップコーディング 2人以上世帯は9000万円以上をトップコーディング	貯蓄減少額： 単独世帯は800万円以上をトップコーディング 2人以上世帯は1300万円以上をトップコーディング	
						借入金残高： 総額以外は提供しない 二人以上世帯は、4500万円以上をトップコーディング 単身世帯は、1500万円をトップコーディング	借入金額： 単独世帯は2400万円以上をトップコーディング 2人以上世帯は4000万円以上をトップコーディング	家計支出総額： 単独世帯は55万円以上をトップコーディング 2人以上世帯は100万円以上をトップコーディング	
		住宅の1か月当たり家賃・間代： 都道府県別に出現頻度により、9万円以上～30万円以上のいずれかでトップコーディング						親への仕送り額： 6万円以上をトップコーディング	
		世帯ごとの家賃・間代： 都道府県別に出現頻度により、9万円以上～30万円以上のいずれかでトップコーディング						子への仕送り額： 16万円以上をトップコーディング	
								育児費用： 7万円以上をトップコーディング	

匿名データの作成に係る匿名化処理基準（案）

資料4別紙2（別添）

	総務省				厚生労働省				
	国勢調査	住宅・土地統計調査	就業構造基本調査	社会生活基礎調査		全国消費実態調査	労働力調査	国民生活基礎調査	
				調査票A	調査票B			匿名データA (世帯票、健康票)	匿名データB (世帯票、健康票、所得票、貯蓄票)
提供項目等 (その4)	<p>世帯主との続柄： 「他の親族」、「住み込みの雇人」及び「その他」をグルーピング</p> <p>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」をグルーピング</p> <p>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</p> <p>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」をグルーピング</p> <p>利用交通手段： 利用交通手段が1種類の場合の「オートバイ」及び「自転車」をグルーピング、「動め先・学校のバス」、「ハイヤー・タクシー」及び「その他」をグルーピング</p> <p>世帯の種類： 「一般世帯」及び「施設等の世帯」の2区分とし、施設等の世帯の内訳は提供しない</p> <p>世帯の家族類型： 公表統計と同じ6区分で提供</p> <p>住居の種類・住宅の所有の関係： 「公営の借家」及び「都市機構・公社の借家」をグルーピング、「給与住宅」及び「間借り」をグルーピング、「住宅以外に住む一般世帯」の内訳（「会社等の独身寮・寄宿舎」及び「その他」）は提供しない</p>	<p>住宅の居住室の畳数： 都道府県別に出現頻度により、5畳未満～7畳未満のいずれかでボトムコーディング、70畳以上～140畳以上のいずれかでトップコーディング</p> <p>世帯ごとの居住室の畳数： 都道府県別に出現頻度により、5畳未満～7畳未満のいずれかでボトムコーディング、70畳以上～140畳以上のいずれかでトップコーディング</p> <p>従前の居住室の畳数： 都道府県別に出現頻度により、5畳未満～7畳未満のいずれかでボトムコーディング、70畳以上～140畳以上のいずれかでトップコーディング</p> <p>住宅の所有名義：提供しない</p> <p>直近5年間の増改築の有無：「東日本大震災による被災箇所の改修工事をした」は提供しない それ以外の項目は提供</p> <p>東日本大震災による転居：提供しない</p> <p>調査票乙のみの項目：提供しない</p>			<p>現在住んでいる住居以外の住宅及び土地に関する項目：提供しない</p>	<p>異動符号：提供しない</p> <p>前月欄の従業上の地位：提供しない</p> <p>前月欄の事業の種類（産業）：提供しない</p> <p>前月欄の勤め先・業主などの企業全体の従業者数：提供しない</p>	<p>同居していない者の状況： 「老人福祉施設入所者」、「社会福祉施設入所者」及び「長期入院者」をグルーピング、「入院・入所」は2人以上をトップコーディング、「単身赴任」と「学業」は3人以上をトップコーディング</p> <p>公的年金・恩給の受給状況： 「福祉年金」及び「恩給」を「その他」にグルーピング</p> <p>在卒の状況と（在卒の）学校の種類： 「在学したことがない」を「卒業」にグルーピング、その場合の「（在卒の）学校の種類」を「小学・中学以下」にリコーディング</p> <p>手助けや見守りを要する者の自立期間： 「1月未満」、「1月～3月未満」、「3月～6月未満」及び「6月～1年未満」を「1年未満」にグルーピング</p> <p>最も気になる傷病： 傷病を上位区分にリコーディング</p> <p>普段の活動ができなかった日数： 25日以上をトップコーディング</p> <p>悩みやストレスの相談状況及び最も気になる悩みやストレスの相談状況：</p>	<p>所得票の世帯員に関する情報：提供しない</p>	

匿名データの作成方針【雛形】(案)

－初めて匿名データの作成を行う統計調査の場合－

1 基本的な考え方

本調査の匿名データ化については、「匿名データの作成・提供に係るガイドライン」(平成 21 年 2 月 17 日総務省政策統括官(統計基準担当)決定)の匿名化処理基準を参考としつつ、本調査の特徴を踏まえて所要の秘匿措置を講じる。

2 作成する匿名データの構成概要

本調査に対する利用ニーズ、調査体系の特性を活かし、世帯又は個人単位の以下の匿名データを作成する。

調査本体の 標本の大きさ	リサンプリング率	匿名データの 標本の大きさ
約〇〇万世帯	〇〇%	約〇万世帯

3 適用する匿名化処理

本調査では、調査票情報に対して、以下の匿名化処理を適用する。

(1) リサンプリング

匿名化処理基準を踏まえ、リサンプリング率〇〇%を目安としてリサンプリングを行う。リサンプリングは、《リサンプリングの方法を記載(例：地域区分による層化を行ったのち、世帯を単位としてまとめた上で等確率抽出により、世帯を単位とするリサンプリング率が約〇〇%になるようにする。)》

(2) 識別情報

以下の秘匿措置を講じるほか、出現数が少なく個体識別リスクが考えられるレコードは削除する。(詳細は、別添「匿名データの作成に係る審査表」を参照)。

例)

ア 世帯人員

世帯人員が多人数である世帯のレコードは削除する。

イ 地域区分

例 1) 地域区分は「全国」のみとする。

例 2) 地域区分は匿名化処理基準を踏まえ「3大都市」と「3大都市以外」の2区分とする。

例 3) 地域区分は都道府県とする。 等

ウ 年齢

個人の年齢は、15歳未満は各歳、15歳以上は5歳階級でグルーピングとするとともに、高齢者についてはしきい値基準に基づきトップコーディングを行う。

エ 住居の床面積

所得総額等は世帯総額のみをトップコーディングし、その内訳情報は削除する。

オ その他

これら以外にも、リスクを低減するために、レコードの削除、トップコーディング、ボトムコーディング、リコーディング、乱数によるレコード順の並び替え等、必要な措置を行う。

また、トップコーディング、ボトムコーディング、リコーディングに当たっては、利便性を考慮すると共に、統計調査の本体集計の結果表章に用いられる分類を参考とする。

匿名データの作成方針【雛形】

－匿名データの年次追加を行う統計調査の場合－

1 基本的な考え方

本調査の匿名データ化については、「匿名データの作成・提供に係るガイドライン」（平成 21 年 2 月 17 日総務省政策統括官（統計基準担当）決定。以下「ガイドライン」という。）の匿名化処理基準に準拠した秘匿措置を講じて作成・提供する。ただし、社会情勢の変化や他調査の作成方法等を勘案し、当該年次の特性に応じた措置を講じる。

2 作成する匿名データの構成概要

今まで作成を行ってきた〇〇〇〇調査に係る匿名データと同様に、以下の匿名データを作成する。

調査本体の 標本の大きさ	リサンプリング率	匿名データの 標本の大きさ
約〇〇〇万世帯	〇〇%	約〇万世帯

3 適用する匿名化処理

本調査では、ガイドラインの匿名化処理基準に準拠した匿名化処理を適用する。

なお、新規の調査項目及び社会情勢の変化等による変更点は以下のとおりである。

(1) 新規の調査項目

- ① 就業期間：50 年以上は 50 年でトップコーディング（1%のしきい値基準）

(2) 社会情勢の変化等

- ① 世帯人員が多い世帯の削除
8 人以上の世帯の割合が減少したため、9 人以上を 8 人以上に引き下げ（しきい値基準）
- ② 年齢のトップコーディング
人口高齢化により 85～89 歳の割合が増加したため、85 歳以上を 90 歳以上に引き上げ（しきい値基準）
- ③ 年収額のトップコーディング
1200 万円以上の割合が減少したため、1200 万円以上を 1100 万円以上に引き下げ（しきい値基準）

(記載例)

匿名データの審査表：記入例

統計調査名： ●●●●	匿名化処理基準	平成●●年 (追加年次)	変更理由・備考	検証結果 (統計研究研修所記入欄)
母集団情報	国勢調査 調査区名簿	同左		
リサンプリング	世帯単位で無作為抽出 (10%)	同左		
調査実施からの経過年数	5年以上	5年		
しきい値	1%	同左		
世帯・個人識別情報の匿名化	世帯人員が9人以上の世帯を削除	世帯人員が8人以上の世帯を削除	世帯人員が8人以上の世帯の割合が低下したため(識別情報の匿名化)	
ノイズやスワッピング処理	なし	同左		
世帯・個人を特定できる外部の情報の有無	なし	同左		
データの並べ替え	同一世帯の世帯員は世帯員番号順を保ったまま世帯員番号順はランダムに並び替え	同左		
提供項目等	○：原則そのまま提供 ▲：匿名化を講じて提供 ×：提供しない			
市区町村番号	×	×	地理的情報	
都市階級区分	○ 人口規模別3区分	○ 同左	地理的情報	
調査区符号	×	×	地理的情報	
調査区の施設等の区分	×	×	特定の種類の施設であることが明らかになる情報	
世帯番号	×	×	地理的情報	
世帯員番号	○	○		
性別	○	○		
年齢	▲ 0～84歳は5歳階級でグルーピング 85歳以上は85歳でトップコーディング	▲ 0～89歳は5歳階級でグルーピング 90歳以上は90歳でトップコーディング	人口高齢化に対応 (識別情報の匿名化)	
住居の床面積	▲ 20㎡以下は20㎡でボトムコーディング 300㎡以上は300㎡でトップコーディング	▲ 同左		
年収額	▲ 1200万円以上は1200万円 でトップコーディング	▲ 1100万円以上は1100万円 でトップコーディング	1%のしきい値基準に基づく。 18年の上位1%は1200万円だったが、 21年は1100万円に減少したため。	
雇用形態	○	○		
職業分類	▲ 「農業作業員」「林業作業員」「漁業作業員」を「農林漁業作業員」にグルーピング	▲ 同左		
就業期間	調査なし	▲ 50年以上は50年で トップコーディング		
乗率	▲	▲	調査区符号が判明する場合は 補正。リサンプリング率も反映	

(注1) 初めて匿名データを作成する統計調査の場合は、「匿名化処理基準」欄は使用しない。

(注2) 記載例を参考に必要な事項を記載し欄を追加する。