

厚生労働省発統 1217 第 1 号
平成 22 年 12 月 17 日

統計委員会 委員長
樋口 美雄 殿

厚生労働大臣
細川 律夫

諮問第 34 号
国民生活基礎調査に係る匿名データの作成について（諮問）

標記について、別紙の通り作成するに当たり、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、統計委員会の意見を求める。

諮 問 の 概 要

(国民生活基礎調査に係る匿名データの作成について)

今回、厚生労働省は、平成 16 年国民生活基礎調査について、統計法（平成 19 年法律第 53 号）第 35 条第 1 項の規定に基づき、以下のとおり匿名データの作成を行う予定である。

1 平成 16 年国民生活基礎調査の匿名データを作成する理由

本調査は当省が実施する統計調査の中でも、保健、医療、福祉、年金、所得等国民生活の基礎的事項を把握しており、学術研究や高等教育においても利用ニーズの高い調査であり、また、世帯対象調査は事業所・企業対象調査より比較的匿名化しやすいことから、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

本調査は、世帯票を中心に、健康票、所得票、貯蓄票及び介護票の 5 種類の調査票で構成されており、利用者における利便性を踏まえて、単一の調査票ではなく複数の調査票を組み合わせた以下の 2 種類の匿名データを作成する。

(1) 世帯票、健康票（世帯単位）：

人口、社会統計分野での世帯数の推計分析等を中心とした利用を想定

(2) 世帯票、健康票、所得票、貯蓄票（世帯単位）：

世帯の所得及び貯蓄に関する分析等を中心とした利用を想定

3 匿名データの作成方法の概要

適用する匿名化措置は、以下の通りである（別添 1 参照）。

(1) 元の統計調査のレコードすべ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（レコードのリサンプリング）。

(2) 直接的な識別情報は、レコードから削除する。また、レコードの配列順が意味をなさないように乱数により並べ替えを行う（識別情報の削除等）。

(3) 特徴的な識別情報の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）。

(4) 極端に大きな（小さな）値は、上限値（下限値）を設けて統合する（トップコーディング、ボトムコーディング）。

(5) 分類事項及び階級は、詳細なものではなく、粗いものとする（リコーディング）。

匿名化技法の概要

一般国民からの信頼と協力により集められた調査票情報を基に作成する匿名データは、学術研究及び教育目的のため広く一般に提供されることから、被調査者が特定できないよう加工することが統計法で規定されている。具体的には、単に氏名・住所を削除しただけでは、年齢、家族構成、職業分類、住居形態等の調査項目と、外観から確認できる情報や一般に入手可能な情報の組み合わせ等により、被調査者が特定されるリスクがあるため、以下の匿名化技法により、被調査者の情報を確実に秘匿するものである。

1 情報の削除

- (1) データの再抽出(リサンプリング)..... 元の統計調査のデータすべてを匿名データに用いるのではなく、そのうちの何割かを再抽出したデータを用いる。
- (2) 直接的な識別情報の削除等..... 直接的な識別情報は、データから削除する。(例:地域を削除)
また、データの配列順により特定されないように、無作為に並べ替えを行う。
- (3) 裾切りによるデータ削除 特徴的で、出現率が低い値があるデータは、削除する
(例:多人数世帯、3つ子以上世帯を削除)

2 識別情報の階級区分統合

- (1) 上限(下限)階級区分の統合..... 極端に大きな(小さな)値は、上限(下限)値を設けて統合する。
(例:一定の年齢以上を統合、所得の一定金額以上を統合)
(トップコーディング、ボトムコーディング)
- (2) 再コード化(リコーディング)..... 分類事項の程度は、詳細なものではなく、粗いものとする
(例:各歳階級を5歳階級化、細かい分類を粗く再分類化)



平成 16 年国民生活基礎調査の概要

1 調査の概要

(1) 調査の目的

本調査は、保健、医療、福祉、年金、所得等国民生活の基礎的事項を調査し、厚生労働省の所掌事務に関する政策の企画及び立案に必要な基礎資料を得るとともに、各種調査の調査客体を抽出するための親標本を設定することを目的とする。

(2) 調査の周期

昭和 61 年以降毎年実施(3 年ごとに大規模調査を実施し、中間年に簡易調査を実施する)。平成 16 年は第 7 回目の大規模調査である。

2 調査の対象及び客体

全国の世帯及び世帯員を対象とし、世帯票及び健康票については、平成 12 年国勢調査区から層化無作為抽出した 5,280 地区内のすべての世帯及び世帯員を、介護票については、同地区から無作為に抽出した 2,500 地区内の要介護者・要支援者を、所得票及び貯蓄票については、前記の 5,280 地区に設定された単位区から無作為に抽出した 2,000 単位区¹内のすべての世帯及び世帯員を客体とした。

3 調査事項

世帯票：性、出生年月、世帯主との続柄、配偶者の有無、公的年金・恩給の受給状況、所得を伴う仕事の有無・就業希望の有無、勤めか自営かの別・仕事の内容(職業分類)、現在の公的年金の加入状況等

健康票：入院・入所の状況、自覚症状、治療の状況、通院・通所の状況、傷病名、支払った費用、日常生活への影響、就床日数、健康意識、悩みやストレスの状況、健康診断受診状況等

介護票：調査票の回答者、介護が必要な者の性別と生年月日、要介護度の状況、介護が必要となった原因、居宅サービスの利用状況等

所得票：所得の種類別金額、課税等の状況、生活意識の状況等

貯蓄票：貯蓄現在高、貯蓄の増減の状況、借入金残高

¹ 「単位区」とは、推計精度の向上、調査員の負担平準化等を図るため、一つの国勢調査区を地理的に分割したものである。

4 標本抽出法

(1) 世帯票・健康票の母集団フレーム、抽出方法等

・母集団フレーム

平成12年国勢調査の調査区のうち、後置番号²が1又は8で、人口が0でない調査区

・層化基準

地域区分：都道府県 / 政令指定都市

(2) 所得票・貯蓄票の母集団フレーム、抽出方法等

・母集団フレーム

第一次：国勢調査の調査区のうち、後置番号が1で、人口が0でない調査区

第二次：世帯票調査区に単位区を設定した名簿

・抽出方法

層化二相二段抽出³

一相目一段目の抽出単位 国勢調査調査区（抽出結果は世帯票調査区）

二相目の抽出単位 国勢調査調査区（抽出結果は所得票・貯蓄票調査単位区を含む世帯票調査区）

二段目の抽出単位 単位区（抽出結果は所得票・貯蓄票調査単位区）

・層化基準

世帯票・健康票と同じ

(3) 介護票の母集団フレーム、抽出方法等

・母集団フレーム

国勢調査調査区のうち、後置番号が1又は8で、人口が0でない調査区

・抽出方法

層化二相抽出

一段目の抽出単位 国勢調査調査区（抽出結果は世帯票調査区）

二相目の抽出単位 国勢調査調査区（抽出結果は介護票調査区）

・層化基準

世帯票・健康票と同じ

² 「後置番号」とは、国勢調査区の種類を表す番号であり、「1」は一般調査区、「8」はおおむね50人以上の単身者が居住している寄宿舍・寮等のある区域をいう。

³ 「二相抽出」とは、母集団の中から一部の抽出単位を親標本として抽出し、情報を調べたうえで、同じ抽出単位で親標本から標本抽出を行う手法である。ただし、ここでの抽出方法は層化二段抽出とほぼ同等なものと考えてよい。なお、大規模年の所得票と貯蓄票の調査対象単位区は同一なので別々に抽出する必要はない。

平成 16 年国民生活基礎調査に係る匿名データの作成方法

1 基本的な考え方

(1) 安全性、国民の信頼の確保

匿名データの作成については、万が一、被調査者が特定、推定されると、本調査のみならず、政府統計全ての信頼が大きく損なわれ、ひいては回収率、記入率の低下を引き起こす恐れがあるため、被調査者の特定、推定の回避はもちろん、一般国民に不安感を与えることのない確実な秘匿措置が必要である。

一方で、匿名データを作成、提供する以上、学術研究及び高等教育における有用性の確保は不可欠であり、秘匿性と有用性はトレードオフの関係にある。

本調査の匿名データ化は、初回であることを踏まえデータの有用性を考慮しつつ、相対的には秘匿性を優先して作成、提供する。

(2) 匿名データ化の基本的方針

本調査の匿名データ化については、本委員会で審議された諮問第 13 号の答申「全国消費実態調査、社会生活基本調査、就業構造基本調査及び住宅・土地統計調査に係る匿名データの作成について」(平成 21 年 3 月 9 日付)による総務省統計局 4 調査の匿名データについて適用されている秘匿措置を参考としつつ、本調査の標本設計、特徴を踏まえて所要の秘匿措置を講じる。

具体的な秘匿措置については、一橋大学を中心とした匿名化技法、人口・社会統計、医学統計等の学識経験者の協力により調査研究いただいた成果を踏まえ、本匿名データの作成方法に反映するものである。

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、世帯又は個人単位で接続可能な複数の調査票情報を接続し、以下の 2 種類の匿名データを作成する(別添 4 参照)。

匿名データの種類の種類	匿名データの構成 (接続する調査票)	リゾプリング率 (提供レポート件数 / 調査本体のリゾプリング)	拡大乗数	想定される利用ニーズ
A	世帯票、健康票	約 2 割 (約 4 万 / 約 22 万世帯)	全国一律の拡大乗数を再付与する。	人口、社会統計分野での世帯数の推計分析等を中心とした利用を想定

B	世帯票、健康票、 所得票、貯蓄票	約2割 (約6千 / 約2万5千世帯)	拡大乗数 は付与し ない。	世帯の所得及び 貯蓄に関する分 析等を中心とし た利用を想定
---	---------------------	------------------------	---------------------	---

3 適用する匿名化技法

特に本調査では、集落抽出により抽出された国勢調査区又は単位区内の世帯を悉皆で調査している特性、複数の調査票情報を接続して匿名化すること等の個体識別リスクを踏まえた総体的な秘匿措置の確保が必要であるため、以下の匿名化技法を適用する。

(1) リサンプリング

前述の集落抽出による集落内世帯悉皆である調査特性及び世帯単位のリサンプリングは情報量が多くなることによるリスク等を考慮し、匿名データA及びBは、国勢調査区(又は単位区)及び世帯の二段抽出により、約2割をリサンプリングする。

(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(別添5参照)。

ア 地域区分

地域区分は、個体識別の可能性を高める最もセンシティブな情報の一つである。
集落抽出である本調査において地域区分を提供することによるリスクを考慮し、秘匿性を確保するため、地域区分は「全国」のみとする。

イ 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。

ウ 個人の年齢

個人の年齢は、原則として5歳階級のグループ化(階級化)を行い、15歳未満は健康票の事項別記入対象年齢等を踏まえた区分とし、上限については、一定年齢でトップコーディング(一律の上限値を与える。)を行う。

エ 同一年齢の子供の数

三つ子以上がいる世帯のレコードは削除する。

オ 父子世帯

父子世帯は削除する。

カ 所得票に関する事項（匿名データBのみ提供）

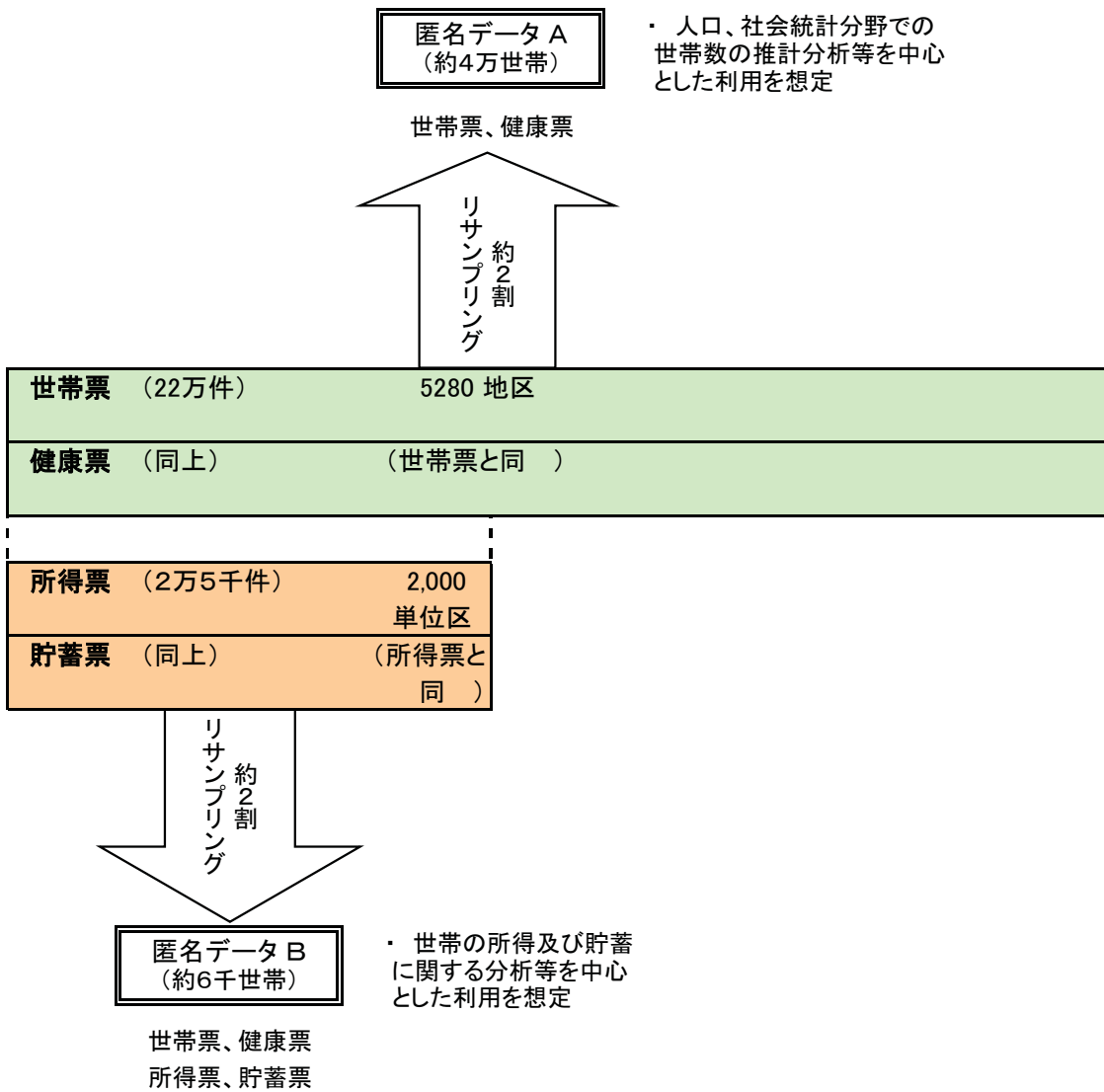
所得総額等は世帯総額のみをトップコーディングし、その内訳情報は削除して提供する。

キ 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（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（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。

平成16年 国民生活基礎調査 匿名データ リサンプリング体系図



平成16年国民生活基礎調査の調査項目と
匿名データの提供項目

提供項目欄の凡例

- :そのまま提供
 ●:匿名化措置を講じて提供
 -:提供しない

調査項目	提供項目
【世帯票】	
<世帯に関する事項>	
地域情報	
都道府県	-
地区番号	-
単位区番号	-
世帯番号	-
世帯人員数	●
世帯構造7分類	●
世帯類型	●
住居の状況	
住居の種類	○
建て方	○
居住室数	●
住宅の床面積	●
単独世帯の区分	○
特定の転出者のいる世帯	●
家計支出総額及び仕送りの状況	
家計支出総額	●
親への仕送りの有無	○
親への仕送り額	●
子への仕送りの有無	○
子への仕送り額	●
所得が最も多い者の世帯員番号	○
乳幼児のいる世帯	
育児費用	●
<世帯員に関する事項>	
世帯員番号	○
世帯主との続柄	○
性	○
出生年月	●
配偶者の有無	○
医療保険の加入状況	○
公的年金・恩給の受給状況	●
手助け見守りの要否	●
要介護認定の有無	●
仕事の有無と就業希望の有無と理由	
仕事の有無	●

調査項目		提供項目
	他の仕事の有無	○
	就業希望の有無	○
	求職活動の有無	○
	希望する仕事の形	●
	すぐには就けない理由	○
	公的年金の加入状況	○
	別居の子の有無	
	別居の子の有無	○
	別居の子の数	●
	最も近くに住んでいる子の居住場所	○
	勤めか自営かの別と仕事の内容(職業分類)	
	勤めか自営かの別	○
	勤め先での呼称	○
	企業規模・官公庁の別	○
	職業分類番号	●
	就業時期・就業期間・通勤時間	
	現在の主な仕事に就いた時期	●
	1週間に仕事をした日数	○
	1週間に仕事をした時間	●
	1日の片道通勤時間	●
	雇用保険の加入状況	○
	乳幼児のいる世帯	
	乳幼児の世帯員番号	-
	乳幼児の日中における保育等の状況	○
	手助けや見守りを要する者の状況	
	世帯員番号	-
	日常生活の自立の状況	-
	手助け、見守りを要する状態になってから期間	-
	主な介護者の状況	
	手助けや見守りを要する者との続柄	-
	同別居の別	-
	性	-
【 健康票 】		
< 世帯に関する事項 >		
地域情報		
	都道府県	-
	地区番号	-
	単位区番号	-
	世帯番号	-
	医療費世帯総額	●
< 世帯員に関する事項 >		
	性	-
	出生年月	-
	入院、入所の有無	○
自覚症状に関する事項		
	自覚症状の有無	○
	自覚症状名	●

調査項目		提供項目
	最も気になる症状(主症状)	●
	主症状の治療状況	○
通院に関する事項		
	通院の有無	○
	傷病名	●
	最も気になる傷病(主傷病)	●
	最も長く通っている傷病	●
	主傷病の通院期間	○
	医療費支出額	-
日常生活影響に関する事項		
	日常生活影響の有無	○
	日常生活影響の事柄	○
	就床日数	●
	健康意識	○
悩みやストレスに関する事項		
	悩みやストレスの有無	○
	悩みやストレスの原因	●
	最も気になる悩みやストレスの原因(主原因)	●
	悩みやストレスの相談状況	●
	主原因の相談状況	●
喫煙に関する事項		
	喫煙の状況	○
	平均喫煙本数	○
健診や人間ドックに関する事項		
	健診受診の有無	○
	健診受診の機会	○
	直近の健診	○
	健診指摘の有無	○
	医療機関受診指導の有無	○
	医療機関受診の有無	○
	健康管理に注意を払うようになったか否か	○
	健診を受けなかった理由	●
	がん検診受診状況	○
【 所得票 】		
<世帯に関する事項>		
地域情報		
	都道府県	-
	地区番号	-
	単位区番号	-
	世帯番号	-
	世帯区分	-
	生活意識	○
	総所得	●
	課税等の状況(税金+社会保険料)	●
	掛金	●
<世帯員に関する事項>		
	性	-

調査項目		提供項目
	出生年月	-
	所得の種類	
	雇用者所得	-
	事業所得	-
	農耕・畜産所得	-
	家内労働所得	-
	財産所得	-
	公的年金・恩給	-
	雇用保険	-
	その他の社会保障給付金	-
	仕送り	-
	企業年金・個人年金等	-
	その他の所得	-
	課税等の状況	
	所得税	-
	住民税	-
	社会保険料	-
	医療保険	-
	年金保険	-
	介護保険	-
	その他(雇用保険等)	-
	固定資産税	-
	掛金	
	企業年金・個人年金等	-
【貯蓄票】		
<世帯に関する項目>		
	地域情報	
	都道府県	-
	地区番号	-
	単位区番号	-
	世帯番号	-
	貯蓄に関する事項	
	貯蓄の有無	○
	金融機関への貯蓄の有無	○
	保険料の有無	○
	株式等の有無	○
	その他の預貯金の有無	○
	貯蓄現在高	●
	貯蓄残高の増減に関する事項	
	貯蓄の増減	○
	減少額	●
	減少理由	○
	借入金に関する事項	
	借入金の有無	○
	借入金額	●