

諮問第 34 号の答申

国民生活基礎調査に係る匿名データの作成について（案）

本委員会は、厚生労働省が作成を予定している平成 16 年国民生活基礎調査（以下「本調査」という。）に係る匿名データの作成方法の計画について審議した結果、下記の結論を得たので答申する。

記

1 計画の適否

本計画については、これにより作成される匿名データにおいて、本調査の調査客体の匿名性及び学術研究等における有用性がおおむね確保されるものと認められることから、適当である。

ただし、以下の「2 理由等」で指摘した事項については、修正が必要である。

2 理由等

(1) 情報の削除

ア レコードのリサンプリング及び地域情報の削除

本調査の匿名データの作成に当たっては、複数の調査票情報を組み合わせることにより、世帯票及び健康票から構成される匿名データ（以下「匿名データ A」という。）並びに世帯票、健康票、所得票及び貯蓄票から構成される匿名データ（以下「匿名データ B」という。）の 2 種類の匿名データを作成することとし、それぞれ、地域区分を「全国」のみとするとともに、国勢調査区（又は単位区）及び世帯の二段階で再抽出（以下「リサンプリング」という。）したものを（以下「サブサンプル」という。）を用いる計画である。

このうち、匿名データ A 及び匿名データ B のそれぞれについて、リサンプリングを国勢調査区（又は単位区）及び世帯の二段階で行うことについては、特に本調査が採用している集落抽出法が、集団を単位として抽出し、抽出された集団内の全ての標本を調査する方法であるという特性を踏まえ、匿名データの中に特定の調査区（又は単位区）が含まれるか否か、また特定の世帯が含まれるか否かの判別を困難とする措置であることから、適当である。

また、地域区分を「全国」のみとし、全国一律の拡大乗数を付与することについては、サブサンプル中に含まれる多くの属性情報と詳細な地域情報を組み合わせた場合に調査客体が特定される可能性が高まること、各レコードが保持する拡大乗数から抽出地域が特定されてしまうことを防ぐ必要があること、本調査の匿名データの作成は今回が初回であり、調査客体の匿名性の確保を十分に図るよう慎重を期す必要があることから、やむを得ない措置である。

以上の措置を講じた結果として、本調査のサブサンプルの抽出率については、約 2 割となっているものの、サブサンプルの大きさは中間年調査の集計客体数と

同程度であること、作成された匿名データA及び匿名データBによる統計と全レコードから作成された公表統計との間で代表的な項目の平均値や分布に大きな乖離はないことから、適当である。

イ 識別情報の削除等

(ア) 直接的な識別情報の削除等

本調査のサブサンプル中のレコードに含まれる情報のうち地区番号等の直接識別できる情報は、これを削除するとともに、レコードは乱数により並び替える計画である。

これらについては、調査客体の特定や探索を防止するために効果的な措置であること等から、適当である。

(イ) 所得の内訳等の削除

所得票に含まれる所得等の情報については、世帯の総所得、課税等の状況及び掛金のみに限定し、その内訳や世帯員別の情報は削除して提供する計画である。

これについては、所得等の内訳や世帯員別の情報を全て提供すると、これらを合計することにより、(2)のアの(イ)による匿名化措置の効果を担保できなくなること、所得等の詳細な内訳の提供は調査客体が特定される可能性を高めることから、やむを得ない措置である。

ウ 裾切りによるレコード削除

本調査のサブサンプル中のレコードのうち、世帯人員8人以上の世帯、同一年齢の子供が3人以上いる世帯、父子世帯、要介護者が2名以上いる世帯、年齢差の大きい夫婦のいる世帯に係るものは、匿名データから削除する計画である。

世帯人員8人以上の世帯のレコードを削除することについては、世帯員の人数の情報が世帯の外部から比較的容易に把握可能な属性であり、それが極端に大きい場合は出現頻度が低く調査客体が特定される可能性が生じること等から、適当である。

同様に、父子世帯、要介護者が2名以上いる世帯、年齢差の大きい夫婦のいる世帯のレコードを削除することについても、外部から比較的容易に把握可能な属性である一方で出現頻度が低く、調査客体が特定される可能性が生じること等から、適当である。

なお、年齢差の大きいまたは小さい親子のいる世帯についても、同様に外部から比較的容易に把握可能な属性であり、調査客体が特定される可能性が生じることから、当該レコードを削除する必要がある。

また、同一年齢の子供が3人以上いる世帯については、本計画では、世帯員の年齢は各歳ではなく年齢階級別に提供されるため、同一年齢階級の世帯員数に着目して、当該階級に4人以上の者がいる世帯についてレコードを削除するよう、変更する必要がある。

(2) 識別情報の階級区分の統合

ア トップコーディング及びボトムコーディング

(ア) 高齢者の年齢

世帯員の年齢については、一定の値を上限値とし、それを上回る場合に上限値以上でまとめる措置（以下「トップコーディング」という。）を行うこととし、当該上限値は85歳以上とする計画である。

これについては、出現頻度が低い一定年齢以上の高齢者をトップコーディングすることにより、性別等の他の属性情報との組み合わせによる調査客体の特定を防ぐことから、適当である。

(イ) 総所得及び貯蓄現在高等

世帯の総所得及び貯蓄現在高等については、一定の金額を上限値としてトップコーディングを行う計画である。

これについては、トップコーディングにより、所得等が極端に大きい調査客体の特定を防ぐことから、適当である。

イ リコーディング（分類区分の再付与）

(ア) 世帯員の年齢

世帯員の年齢（トップコーディングを行う高齢者を除く。）については、その分類の程度を粗いものにする措置（以下「リコーディング」という。）を講じることとし、15歳以上の者は5歳階級別とし、15歳未満の者は「0～5歳」「6～11歳」及び「12～14歳」の3区分とする計画である。

これについては、各歳別のデータ提供に比べて匿名データの有用性が低下するものの、各歳別の年齢が明らかになると、世帯員に関する他の多くの属性情報との組み合わせにより調査客体が特定される可能性が生じることから、やむを得ない措置である。

なお、15歳未満の者については、健康票の記入対象項目が年齢により異なることから、年齢の特定を防ぐために健康票の回答区分である3区分としたものであり、やむを得ない措置である。

(イ) 出現頻度の低い選択肢のある項目

出現頻度の低い選択肢のある項目については、当該選択肢を「その他」等に統合する計画である。

これについては、調査客体の特定を防ぐことから適当な措置であるが、「希望する仕事の形」、「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）」及び「健診を受けなかった理由」については、専ら本人の意識を問う項目であって外観から識別される可能性が低く、当該情報によって調査客体が特定される可能性が低いと考えられることから、匿名化措置を緩和し、匿名データの有用性の向上を図る必要がある。

ウ トップコーディング等の基準

トップコーディング、一定の値を下限值としこれを下回る場合に下限値以下でまとめる措置（以下「ボトムコーディング」という。）等を講じる場合、本調査では対象サンプル全体の1%未満を対象とする計画である。

これについては、本調査が集落抽出法で実施され、各世帯及び世帯員に関する多様な項目が把握されていること、有用性の観点から閾値は可能な限り継続した方が望ましく、他年次の匿名データの作成においても当該閾値により匿名性が確保されることを考慮したものであり、やむを得ない措置である。

3 今後の課題

本計画については、本調査に係る匿名データの作成は初回であって、多様な調査項目や抽出方法を考慮した場合、調査客体の匿名性の確保により慎重を期する必要があることから、厳格な匿名化措置を講じていることはやむを得ない。

しかしながら、匿名データの利用者のニーズについては様々なものが考えられることから、以下の課題等について速やかに検討を進め、当該データのより一層の充実に努める必要がある。

（1）地域区分及びリサンプリングの単位

本計画では、匿名性を確保するため、調査客体である世帯の特定につながる可能性が高い地域情報を削除し、地域区分を「全国」のみとする厳格な匿名化措置を講じていることとしている。

しかしながら、地域区分については、有用性の観点から極めて重要な情報であることから、調査客体の匿名性の確保を十分に図りつつ、匿名データの利用者のニーズを踏まえて、何らかの地域表章の可能性について検討する必要がある。

また、リサンプリングの単位については、今回、世帯単位のみとしているが、世帯員単位でリサンプリングを行うことで地域情報の付与やリサンプリング率の向上の可能性があると、公衆衛生や疫学分野の研究においては、世帯員単位での健康状態や生活習慣の分析が重要となること等から、利用者のニーズを十分に考慮したうえで、世帯員単位でのリサンプリングによる匿名データの作成の可能性について、速やかに検討を開始する必要がある。

（2）所得票の情報の提供

本計画では、所得票に含まれる情報については、世帯の総所得、課税等の状況及び掛金のみに限定して提供することとしている。

しかしながら、近年、社会保障や所得格差等に関する研究の重要性が増しており、その分析には所得等に関する内訳や世帯員別の情報が重要であること、一方、本計画で適用されていないトップコーディング等以外の匿名化措置の適用も考えられることから、今後、匿名化措置に関する研究等の進展や利用者のニーズを十分に考慮したうえで、所得等の内訳や世帯員別の情報の提供の可能性について検討する必要がある。

(3) 匿名データの作成対象年次の拡大

本計画では、匿名データの作成対象調査を調査実施後5年以上経過したものとしており、今回は平成16年に実施したものを作成対象とするとともに、今後、順次拡大することとしている。

しかしながら、研究には経年的な分析が重要であるとともに、近年の経済・社会状況の急激な変化に伴い直近の統計に基づく分析の重要性が増していること、さらに、本調査については3年ごとに大規模調査が実施されていることを踏まえれば、提供時期の短縮について検討する必要がある。

(4) 年齢のトップコーディング

本計画では、世帯員の年齢については、85歳以上でトップコーディングを行うこととしている。

しかし、トップコーディングの上限値については、近年の急速な高齢化の進展及び高齢者に関する分析の重要性等を踏まえ、今後、匿名データの作成対象年次を拡大する際には、当該年次の人口構成に応じて検討する必要がある。

(5) トップコーディング等が行われた変数

本計画により作成された匿名データの各レコード上の変数のうち、トップコーディング及びボトムコーディングが行われている変数については、利用者の利便性向上の観点から、海外における提供事例も踏まえ、当該トップコーディング等を行った変数の基本統計量等の提供可能性を速やかに検討する必要がある。

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）	リサンプル率（提供レコード件数／調査本体のサンプルサイズ）	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）

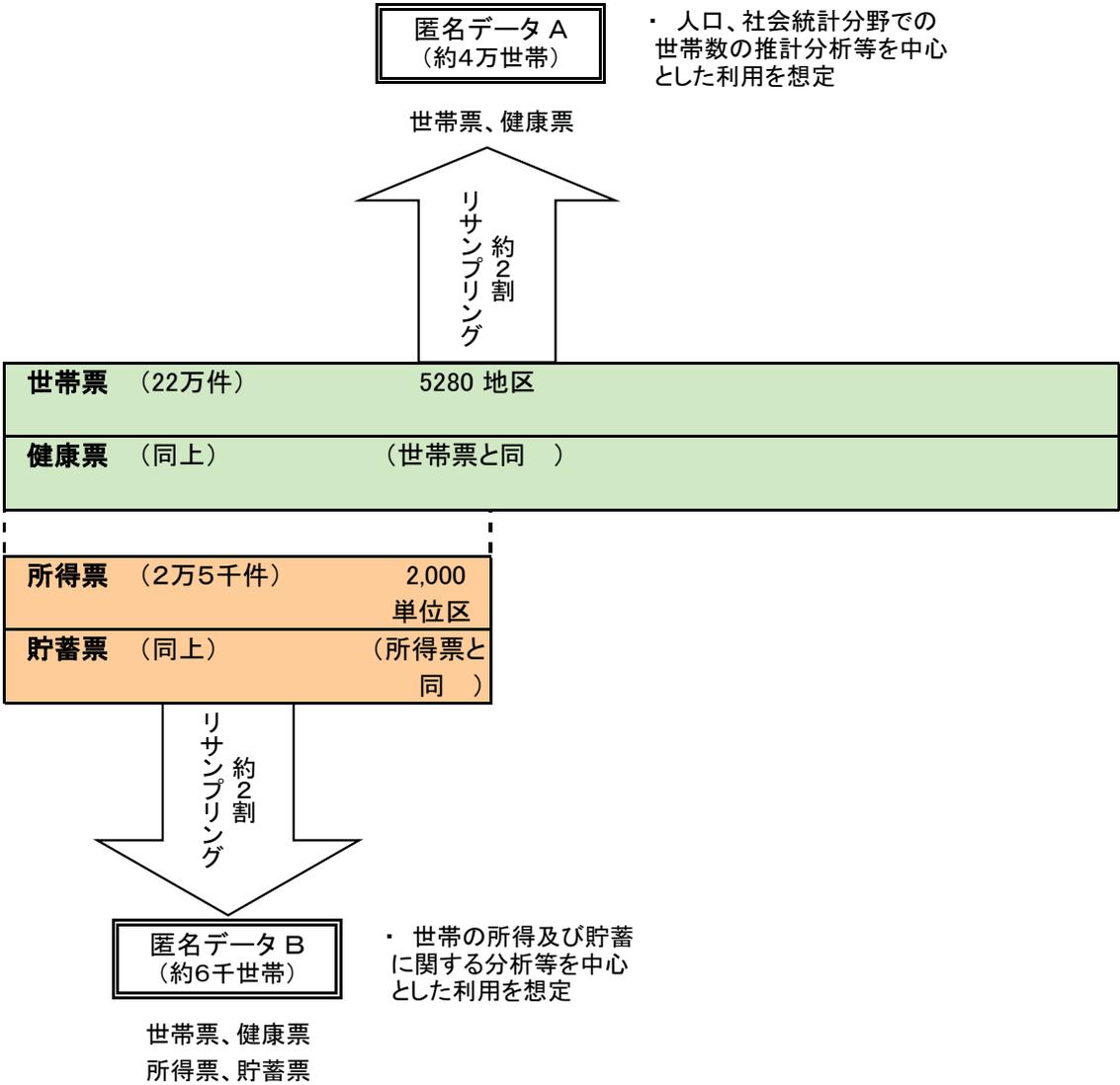
所得総額等は世帯総額のみをトップコーディングし、その内訳情報は削除して提供する。

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	●

第6回匿名データ部会 議事概要

1 日 時 平成23年3月8日(火) 14:58~16:47

2 場 所 中央合同庁舎第4号館2階 共用第3特別会議室

3 出席者

椿広計部会長、井伊雅子部会長代理、津谷典子委員、廣松毅委員、伊藤伸介専門委員、安田聖専門委員、石井太氏(国立社会保障・人口問題研究所)、総務省(統計局)、農林水産省、経済産業省、国土交通省、日本銀行、東京都、千葉県

【諮問者(厚生労働省統計情報部)】

中島企画課審査解析室長、山田企画課審査解析室室長補佐、久住企画課審査解析室匿名データ提供係長

【事務局(内閣府統計委員会担当室)】

若林参事官、谷道参事官補佐

4 議事次第 (1) 国民生活基礎調査に係る匿名データの作成について
(2) その他

5 議事概要

(1) 第43回統計委員会における統計委員会委員の意見について

事務局から、資料2「第43回統計委員会における統計委員会委員の意見」が紹介された。

(2) 前回部会での指摘を踏まえた秘匿措置の見直しについて

諮問者から、前回の部会審議で指摘された事項に対して、検討の結果、秘匿措置を見直したい旨の回答があり、委員から当該回答に対する意見等が述べられた。

各委員等の主な意見等は次のとおり。

厚生労働省の回答

- ・ 裾切りによるレコードの削除のうち、「年齢差の大きい親子」の含まれる世帯に加えて「年齢差の小さい親子」の含まれる世帯も削除することとしたい、「同一年齢の子供が3人以上いる世帯」について、世帯員の年齢が年齢階級別に提供されることを踏まえ、「同一年齢階級に4人以上いる世帯」のレコードの削除へ変更することとしたい。
- ・ リコーディングを予定していた項目のうち、世帯票の「希望する仕事の形」、健康票の「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(主要因)」及び「健診を受けなかった理由」について、専ら意識を問う項目であり、外観から識別される可能性が低いことから、そのまま提供することとしたい。

委員等の意見

- ・ 「年齢差の大きい親子」について、10階級以上離れている場合はレコードを削除することとしてい

るが、父親と子の年齢差に関しては10階級離れていてもそれほど稀な事例とは言い切れない。父親と母親のそれぞれで、子との年齢差を確認すべきではないか。

部会長のまとめ

- ・ 「年齢差の小さい親子」の世帯のレコードを削除の対象とすることは適当。なお、「年齢差の大きいまたは小さい親子」の世帯については、父親や母親で層別に見て0.5%を超えるような時に秘匿措置の緩和が可能であるか、実施者には確認していただき、結果については部会長一任としたい。
- ・ 「同一年齢の子供が3人以上いる世帯」について、「同一年齢階級に4人以上いる世帯」の削除に変更することは適当。
- ・ 「希望する仕事の形」以外に、「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(主要因)」及び「健診を受けなかった理由」に係る秘匿措置を緩和することは、有用性の観点から適当。

(3) 答申(案)の審議について

部会長から、資料1「諮問第34号の答申『国民生活基礎調査に係る匿名データの作成について(案)』」が示され、事務局が答申(案)について項目ごとに朗読の後、審議が行われた結果、一部で所要の修正を行うこととされたが、答申案は概ね適当であるとして採択された。

なお、答申(案)の修正文の表現については部会長に一任することとされた。
各委員等の主な意見等は次のとおり。

ア 裾切りによるレコード削除(2の(1)のウ)

委員等の意見

- ・ 当初の計画になかった「年齢差の大きいまたは小さい親子」について、部会審議では「年齢差の大きい親子」を提供することは問題がないかとの指摘があったが、実施者による検討の結果、「年齢差の小さい親子」についても問題が見つかったので、こういった部会の審議経過を正確に記述する意味で、案2が適当ではないか。

部会長のまとめ

- ・ 「年齢差の大きいまたは小さい親子」については、案2を採用することとしたい。
- ・ 同一年齢階級に一定以上の者がいる世帯について、レコードを削除することについては、「一定以上」ではなく「4人以上」と明確にすべき。

イ 出現頻度の低い選択肢のある項目(2の(2)のイの(イ))

委員等の意見

- ・ 「専ら本人の意識を問う項目であって」の部分に対し、削除の意見と「本人の意識に関わる項目でもあって」という修正意見の2つの意見があるが、この部分は本人の意識上でしかわからないから緩和できるという意図があり、単純に外観識別可能性が低いというだけでは問題。
- ・ 「意識を問う」となると一種のアンケート調査、意識調査というような位置づけになるように受け取られる可能性があるので「意識に関わる」としてはどうか。
- ・ 「本人の意識に関わる」とすると、他にも意識に関わる項目があるが、それは緩和できないので、「意識を問う」としたものである。

部会長のまとめ

- ・ 「専ら本人の意識を問う項目であって」については、原案通りとすることとしたい。

ウ トップコーディング等が行われた変数（3の（5））

厚生労働省の回答

- ・ トップコーディング等が行われた変数の平均値等の提供については、まだ提供実績もなく統一的・基本的考え方も整理されていないこと、該当世帯数が少ない場合に秘匿措置として十分かどうかやサンプルによってバラツキが非常に大きいなどの検討すべき事項があることが判明したことから、直ちには提供できず、検討の時間が必要。

総務省の意見

- ・ 総務省の4調査について、技術的にはどのくらい詳細な区分で出すのか、また、該当する対象が少ない場合は秘匿が破られる危険性があるという問題があるが、ご要望のあったデータを出せるよう検討したい。

委員等の意見

- ・ 公表統計と違った結果が出たときに、それが論文としてアクセプトされるのかという心配がある。今後、分野によってはあるかもしれないが、総務省4調査のときには当該平均値等がないから論文がアクセプトされなかったということは聞いたことはない。本来はオーダーメード集計で集計すべき情報ではないか。
- ・ トップコーディング等が行われた変数の平均値等については、メタデータという形で提供する方法もあるが、対象サンプルのレコードの中にトップ（ボトム）コーディングの閾値ではなく、当該部分の平均値等を入れるという方法も考えられるので、海外の提供事例も含めて検討する必要がある。
- ・ オーダーメード集計でできる場合もあるだろうが、匿名データとして利用する際に、トップコーディングした部分をインピュートしようとしてもどのような値を入れればよいのかわからないので使い勝手が悪いという意味であるならば、現時点では結論は出せず、更に検討が必要。
- ・ 対象の数がどの程度であればよいか、どのような情報の開示の仕方がよいのかについて技術的検討が必要である。ただ、匿名データの有用性の観点からすると、開示のリスクがそれほどないのであれば、その部分の補助情報を提供することは有用性があると思う。
- ・ トップコーディングした部分の平均値はかなり不安定になっている。サブサンプル数だけでなく、リサンプリングによるところも大きい。このまま公表されるのはむしろ問題であり、ミスリードされ、公表統計と違った結論を出されることを危惧する。
- ・ トップコーディングした部分の平均値を出すことは極めてミスリードを起こす可能性があることに注意した方がよいが、トップコーディングした変数全体の基本統計量であれば出せるのではないか。総務省4調査を含めて、統一的な基本統計量を出すことについての情報開示の方法論を検討していただきたい。

部会長のまとめ

- ・ 「トップコーディング等を行った変数の平均値等の提供可能性を速やかに検討する」の「平均値」を「基本統計量」に変更する。
- ・ この点については、総務省4調査のときも今後の課題となっていたので統一的に検討していただきたい。また、今後、他の基幹統計調査において匿名データを作成する際にも同じ問題が発生するため、基本統計量を提供することについて恒常的に配慮していただく必要があるということ、次の統計委員会において答申案を説明する際に、資料2の回答として報告したい。

エ その他の意見について

委員等の意見

- ・ 「3 今後の課題」に関して、「(1) 地域区分及びリサンプリングの単位」及び「(2) 所得票の情報の提供」の中で、潜在的な利用者のニーズにも配慮されるように「利用者のニーズを十分に考慮したうえで」と修文した。また、「匿名データA及び匿名データBの閾値」については、今後の課題に書くには少し技術的すぎる指摘ということで削除することとしたい。
- ・ 所得等の内訳の提供についても、総務省4調査とも連携を図って検討していただきたい。
- ・ 「調査客体」という言葉が多いが、「調査対象」とか別の表現を使っている場合もある。他の答申でどのように表現しているか見て検討してほしい。

以上

<文責 内閣府大臣官房統計委員会担当室 速報のため事後修正の可能性あり>