

匿名データ部会の審議状況について（報告）

第4回匿名データ部会 議事概要

1 日 時 平成22年12月21日(火) 10:00~12:01

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、宇賀克也委員、伊藤伸介専門委員、黒田祥子専門委員、橋本英樹専門委員、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、東京都、千葉県、日本銀行

【諮問者(厚生労働省統計情報部)】

中島企画課審査解析室長、上田社会統計課国民生活基礎調査室長、山田企画課審査解析室課長補佐、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)匿名データ部会の運営等について

(2)国民生活基礎調査に係る匿名データの作成について

(3)その他

5 議事概要

冒頭、部会長、委員及び専門委員、審議協力者並びに事務局のあいさつに引き続き、井伊委員が部会長代理に指名された後、次の議事が進められた。

(1)匿名データ部会の公開について

事務局から、参考1、参考2及び参考4に基づき、匿名データ部会の公開方法等について説明があり、調査客体の特定リスクの防止の観点から、会議及び議事録は非公開に、また、議事概要及び配布資料は公開にすることで確認された。

(2)国民生活基礎調査に係る匿名データの作成について

まず、調査実施者から資料2に基づき、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」の内容が説明され、続いて、事務局から資料3に基づき、本諮問に関する第41回統計委員会での委員からの意見が紹介された。

更に資料2及び資料3の説明を踏まえ、部会長から資料4に基づき、当該作成に関する論点(案)が示され、個別の論点に沿って審議が行われた。各委員等の主な意見は次のとおり。

ア リサンプリングの単位や割合等に関する意見

- ・ 本調査は集落抽出法という標本抽出を行っており、開示リスクを十分に低くするためには、地域ごとの乗率を考慮したリサンプリングを行う必要がある。必ずしも初めから2割を目途に抽出しようとしたわけではなく、二段抽出した結果としてリサンプリング率が2割となっているものであり、

世帯単位での抽出を前提とする限り、やむを得ない措置であろう。

- ・ 本調査が、世帯の構成、所得、健康状態など世帯に関する詳細な情報を収集していることを鑑みると、今回、世帯単位での匿名データが作成されることは、人口や社会保障の分野の分析者にとって非常に有用なデータ提供になることを評価したい。
 - ・ 全国一律の抽出率とするために一番抽出率の低い県（乗率の大きい県）に合わせてリサンプリングしたことについては、地域情報を秘匿する観点及び全国ベースでの推定を行う観点から見て適当であると考ええる。
 - ・ リサンプリング率を議論する時には、地域区分との関係と併せて議論すべきであり、乗率から地域が特定されるリスクがあるというのであれば、今回、地域区分を全国一本というような形で二段抽出したことは理解できる。
 - ・ 公衆衛生や疫学分野の利用者ニーズでは、世帯単位よりもむしろ個人の健康状態や生活習慣などが世帯要因とどう関係しているかといったことが分析できるような匿名データを作成していただきたい。安全性の面からみても世帯単位での提供よりも世帯員（個人）単位での提供の方がはるかに安全であるので、必ずしも世帯レベルにこだわることなく、世帯員レベルでの匿名データの作成を検討いただきたい。また、併せて、世帯単位と世帯員（個人）単位の匿名データを2つ作成するなど、複数の匿名データを作成することの可能性についても検討していただきたい。
- 今回の諮問内容にあるリサンプリングの方法論自体は概ね理解が得られたものの、リサンプリング方法の適否はそれ単独で議論するのではなく、地域区分などとの関係で議論する必要がある。
- また、諮問された内容には無いものの、世帯員レベルでの匿名データの作成になどについては、将来に向けた有用性の観点で審議していきたい。

イ 地域区分に関する意見

- ・ 地域区分については、地域ごとの物価水準を踏まえた実質的な購買力を見ていきたいところ。都道府県別で提供されればありがたいが、せめて地域ブロック単位での情報を提供していただきたい。
 - ・ 総務省統計局4調査の時に採用した都市部とそれ以外といった地域区分も検討していただきたい。
 - ・ 地域情報について、標本設計の段階で都市部とそれ以外といった設計がされていないので、当該区分での提供は困難。また、都道府県単位やブロック別という区分での提供は可能だが、都道府県単位で付与されている乗率に大きな差があるため、地域区分を付与した形でリサンプリングするとデータ数が少なくなる可能性がある。
- 今回は安全性を優先して全国一本での提供としているが、地域情報を付加した情報を利用したいのであれば、統計法第33条による調査票情報の提供で対応すべきではないか。
- 地域区分について、厚生労働省には次回の部会までに「都市部とそれ以外」や「ブロック別」といった区分の地域情報を付与した場合のリサンプリング率について、可能な範囲で結構なので検討

していただきたい。

ウ その他の秘匿措置に関する意見

- ・ 総務省統計局の4調査では、全体に占める割合が0.5%未満のレコードについて裾切りによるレコード削除を施したが、本調査では年次によりリサンプリング率が変わる可能性があるため、やや余裕を持って1%基準で裾切りを行っているが、経年的な分析をするうえでは妥当な措置と考える。
- ・ 今回の匿名データについて、ユーザーの立場からは有用性よりもやや安全性に偏った提案に思われるので、地域情報を一切提供しないのであれば、世帯員の年齢区分をもう少し詳細にしたり、高齢者の健康状態を分析できるようトップコーディングを緩和したりできないか検討していただきたい。
- ・ 例えば、1週間の労働時間は80時間以上でトップコーディングしているが、これらの情報と健康状態をクロスさせて分析することは大変有用。個人的には個人の労働時間が外観識別可能な情報とは思えないので、どういった場合にトップコーディングしなければならないのか、しなくてもよい場合はあるのかなどについて議論していただきたい。
- ・ 匿名データは、学术研究や高等教育を目的とする者に利用を制限しているので、一般国民すべてが目にする可能性は非常に低い。現状では匿名データは個票データと同じ扱いになっているので、多少扱いがきつくなっている。
- ・ 匿名データA及びBのトップコーディングが、公表統計の全サンプルにおける出現率を基にしているのであれば、所得票・貯蓄票の調査対象からのみリサンプリングされている匿名データBの分布が公表統計の分布とは違う可能性もある。その場合、匿名データBにおいて世帯票・健康票に含まれる特定の量的項目に関しては、所得票・貯蓄票の対象サンプルにおける分布に基づいてトップコーディングを適用するほうが望ましいのではないかと考える。

エ その他の意見

- ・ 総務省統計局の4調査に係る匿名データ部会の審議では、安全性と有用性の2点を基準に審査されていたが、今回、「国民感情への配慮」といった新しい基準により匿名データの作成が行われている。本調査の匿名データ作成に当たっては新しい基準も踏まえて審議するのか、判断基準を明確にしたい。
- ・ 一般に行政記録情報を開示する場合、いわゆるモザイク・アプローチによって特定の個人の識別ができるかどうかの問題となるが、調査票情報の場合では、いわゆる一般人が入手し得る情報での照合が可能な場合（一般人基準）と、特定人が持っている情報との関係で照合が可能な場合（特定人基準）のいずれの基準で他の情報との照合可能性を考えるのか。また、他者からは十分に匿名性が確保されていても本人には分かる場合、そのことが調査協力にマイナスの影響が出るかもしれないので、その辺りをどう担保するのか。
- ・ 所得票のデータ等を用いてジニ係数などを分析したいと考えた場合、所得等の平均値が公表統計

と匿名データの間で整合的であるだけでなく、そのばらつきなどについても整合的であることが重要なので、所得に関するばらつきが公表統計との分布状況なども確認すべき。

→ 厚労省には次回の部会までに所得の分布状況が分かる資料を提出するようお願いする。

- ・ 介護票については、調査対象となる世帯の出現率が非常に低いこと、要介護者・要支援者は外観識別性が非常に大きいこと、リサンプリングにより精度が維持できないことから有用性に欠けてしまうこと、などの理由から匿名データ化をしないという説明があったが、次回以降これに関してご意見があれば議論したい。

(3) その他

部会長から、本部会の審議内容の特殊性に鑑みて、総務省統計局の4調査の匿名データの作成に係る部会審議（平成21年1月～3月）に参画していただいた津谷委員と廣松委員を、次回以降の部会審議に追加することについての提案があり、出席の委員及び専門委員からの了解を得られた。

このため、1月26日（水）の第42回統計委員会において、第4回匿名データ部会の審議報告を行う際、本件についても併せて部会長から提案することとなった。

また、次回の匿名データ部会は2月7日（月）に開催することとなった。

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>