

第11回 匿名データ部会 議事概要

1 日 時 平成 24 年 12 月 7 日 (金) 10:00 ~ 12:02

2 場 所 中央合同庁舎 4 号館 2 階 共用第 3 特別会議室

3 出席者

(部 会 長) 椿広計

(委 員) 安部由起子、津谷典子

(専 門 委 員) 伊藤伸介、加藤久和、安田聖

(審 議 協 力 者) 総務省 (政策統括官 (統計基準担当))、財務省、文部科学省、厚生労働省、農林水産省、経済産業省、国土交通省、東京都

(諮 問 者) 総務省統計局統計調査部 : 井上調査企画課長、岩佐国勢統計課長ほか

(事 務 局) 内閣府統計委員会担当室 : 村上室長、若林参事官ほか

4 議 事

(1) 国勢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) 国勢調査に係る匿名データの作成について

冒頭、事務局から前回の部会審議等を踏まえて修正された資料「『国勢調査に係る匿名データの作成について』の論点」の説明が行われた。その後、論点の項目ごとに、諮問者から論点への回答について説明を受けながら審議が行われた。各委員等の主な意見等は次のとおり。

ア 作成に使用するデータについて

- ・ 調査票に記入された内容を数値や符号に置き換えた全数データの電磁的記録を用いることは、適当と判断する。

イ 地域区分について

- ・ 地域区分を「都道府県」及び「人口 50 万以上の市区」とすることは、適当と判断する。

ウ サンプルングの方法等について

- ・ サンプルングの方法には異論はないが、検証に用いている地域の理由を伺いたい。
国勢調査はデータの量が多く、全ての地域のデータを示すことが困難であることから、都市や郊外等の多様な地域特性を持つものと考えられる地域を検証対象として示したものである。なお、検証に当たっては、この地域以外の地域も用いている。
- ・ 前回、「一般世帯」と「施設等の世帯」で抽出単位が異なっていたため、ウエイト付与の可能性について議論になったと記憶しているが、抽出単位が異なることによる影響はどの程度か。
抽出方法上、人口に対する影響はないものとなるが、世帯については、施設等の世帯は個人単位で抽出することから、これにより、見かけ上、単身世帯が増加することになるが、施設等の人口は

少なく、世帯全体への影響は軽微である。また、施設等の世帯はそれ用のフラグを設けており、元々の単身世帯との区別は容易に判るものとなっている。

- ・ 全体の数値にほとんど影響がなく、施設等の世帯と単身世帯が区別できるようにフラグが付与されているということなので、ウエイトを付与する必要はないのではないかと考える。
- ・ 全体の平均や分布には影響がないとしても、ウエイトなしで集計した場合、クロス表では異常値が出る可能性があるのではないか。

どのクロス表の項目を基準としてウエイトを作成するのかという問題があり、任意のクロス表の精度を全て保証することは難しいと考える。むしろ利用者が目的に応じ、国勢調査の公表結果などを基に一定の基準値を設け、ウエイトを算出する方がよいと思われる。

- ・ 利用者が個々の分析に応じてウエイトを計算して調整すればよいのではないか。様々な場合を考慮してウエイトを作成することについては、費用対効果を考えると、費用の方が高くなると思う。
- ・ レコードの削除も考慮したウエイトの作成には時間がかかるので、今回の審議の中でウエイトの付与について具体的に検討するのは難しいのではないか。今後、利用者からのニーズが強まれば、将来的な検討課題になり得る。
- ・ 利用者は1%抽出であることを理解して利用すると考えられるので、費用対効果を考えると、ウエイトは付与しなくてもよいと考える。
- ・ 匿名データによって作成したクロス表と公表統計との乖離については、それらが比較できるような情報を利用者に提供し、その情報を踏まえて利用していただく方がよいと思う。
- ・ 利用者が公表値からウエイトを算出できるので、作成者の方でウエイトをデータに付与する必要はないと考えられる。

部会長のまとめ

- ・ 全世帯を母集団として1%をサンプリングすることとし、「一般世帯」については世帯単位で、「施設等の世帯」については個人単位で抽出することについては、分布がそれほど影響を受けていないという観点から、適当と判断する。
- ・ 「施設等の世帯」と「一般世帯」の抽出単位が異なることに関して、利用者に対する注意事項を出すということと、代表的な公表値について、匿名データにより作成したクロス表と公表されている結果とを比較できるような情報を利用者に提供するというものをもって、ウエイトを付与しないことについては適当と判断する。

エ スワッピングの導入について

- ・ スワッピングの導入自体は問題がないと思うが、利用者への説明を考える必要がある。
- ・ スワッピングの導入及び方法については、有用性の観点からも、原データとの誤差的な検討を踏まえて確認したところ、あまり大きな差はなかったため、適当と判断する。利用者への説明の方法については、次回以降に議論することとしたい。

オ 情報の削除について

(ア) 直接的な識別情報の削除

- ・ 人口 50 万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、調査客体を直接識別できる情報を削除することは、適当と判断する。

(イ) 出現頻度が低い又は特徴的な値があるレコードを含む世帯の削除

世帯人員が多い世帯

- ・ 世帯人員が多い世帯を削除した場合の影響は、年齢階級別にみると違ってくるということか。
世帯人員が多い世帯は、3世代世帯などと重なることになると考えられることから、その影響が幼年層や高齢者層に若干出てくるのはやむを得ない。

部会長のまとめ

- ・ 地域区分ごとに削除の基準を変えて、世帯人員が7人以上～9人以上いる世帯を削除することについては、世帯単位でデータを提供しており、トップコーディング等の他の匿名化措置では対応できないことから、適当と判断する。

父子世帯

- ・ 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を削除することについては、クロス集計をかけなくとも、全ての地域区分において、一般世帯に占める父子世帯の割合が非常に小さく、0.5%基準を大幅に下回っているため、匿名性の観点からは削除しないで提供することは困難であり、適当と判断する。

年齢差の大きい夫婦のいる世帯、年齢差の大きい又は小さい親子のいる世帯、及び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い世帯

- ・ 今回、クロス表の分布をみて削除を行うという、従来の0.5%基準とは異なる方法を適用することは、年齢差の大きい夫婦のいる世帯といった外観識別性が高いと考えられる属性をさらにクロスをしたときに出てくるような母集団一意を避ける方法であると理解しており、問題ないと考えられる。また有用性の観点からは、あまり基準を高くし過ぎない方が、削除する数が多くならないのでよいと考えられる。
- ・ 基本的にはこれでよいと思うが、年齢差の大きい夫婦や子供等のいる世帯について、0.5%基準による方法と今回の方法では、それぞれの程度のレコードが削除されるのかについて、利用者の有用性の観点から確認したいので、可能な範囲で資料を準備いただきたい。

次回以降、資料を提示する。

部会長のまとめ

- ・ 年齢差が25歳以上の夫婦のいる世帯を削除すること、年齢差が45歳以上の親と子、年齢差が14歳以下の親と長子又は19歳以下の親と末子のいる世帯を削除すること、及び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を削除することについては、適当と判断するが、有用性の観点から、削除の状況について確認することとしたい。

(ウ) 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を含む世帯の削除

- ・ 例えば、公表されている市区町村レベルの小地域集計表で、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についても、削除するのか。

全数調査特有の処理として、既存統計表からの特定を防ぐための削除であることから、匿名データの提供地域と同じ単位の報告書が対象となるため、市区町村単位の表は対象とはならない。

- ・ 匿名データの作成手順について確認したいが、母集団一意又は二意のレコードを含む世帯を削除してからスワッピングを行うのか。

そのような手順だったと思うが、確認したい。

部会長のまとめ

- ・ 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を削除すること、及び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項目が含まれる統計表で、地域において、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を削除することについては、適当と判断する。

カ 世帯員及び世帯に関する項目の再編等について

(ア) 年齢

- ・ 平成 12 年及び平成 17 年国勢調査では、現在と比べると 85 歳以上の人口は少ないので、85 歳以上をトップコーディングすることはやむを得ないと考えられるが、今後急激に高齢化が進むことを考慮すると、今後の匿名データの提供の際には検討が必要である。
- ・ 雇用労働力化が進んでいるので、産業、就業状態、配偶関係等をクロス集計すると、必ず少ないところが出てきてしまう。この場合、85 歳以上は必ずトップコーディングをすることになってしまうので、これをどう考えていくのかということは、今後検討していくべき課題であると思う。
- ・ 個人ベースとか、地域区分なしのデータであれば、提供する項目を増やすことはできるのではないかと。項目を削除して提供するというのも、今後の検討課題ではないか。

部会長のまとめ

- ・ 各歳別のクロス集計では、母集団一意、二意が頻出していることから、0～84 歳を 5 歳階級区分とすること及び 85 歳以上をトップコーディングすることについては、有用性という観点では問題があるが、今回はやむを得ない措置であるということで、まとめることとしたい。

(イ) 世帯主との続き柄、国籍、5 年前の住居の所在地（平成 12 年国勢調査）労働力状態、就業時間、従業上の地位

- ・ 地域区分については、「都道府県」及び「人口 50 万以上の市区」を提供しているため、そのトレードオフとして、他の属性に関して分類区分の統合を行うことは、やむを得ないものと考えられる。諸外国でも同様の考え方に基いてマイクロデータの提供が行われている。
- ・ 従業上の地位については、例えば年齢階級によって「雇人のある業主」などの分布に違いがあるが、年齢階級別に統合の基準を変えることはあり得るのか。

年齢階級によって統合の仕方を変えるとすると、その年齢階級のレコードの「従業上の地位」の表示の仕方が他の年齢階級と異なることになるなど、データ操作の際の利用者の負担が大きくなることから、統合の基準は一律の方がよいと考える。

部会長のまとめ

- ・ 母集団一意などが、例えば特定の年齢層に偏っているような場合に、統合の基準を分布に基づいて変えるというような方法については、今後の検討課題として整理したい。世帯主との続き柄、国籍、5 年前の住居の所在地、労働力状態、就業時間、従業上の地位の秘匿の方法については、適当と判断する。

(2) その他

- ・ 次回の匿名データ部会は、12 月 25 日（火）13 時から中央合同庁舎第 4 号館 共用第 2 特別会議室で開催することとされた。また、本日審議できなかった残りの論点については、次回以降に審議することとされた。
- ・ また、予備回としていた平成 25 年 1 月 31 日（木）に第 13 回会合を開催し、答申案の取りまとめを行うこととされた。

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>