

## 匿名データ部会の審議状況について（報告）



## 第12回 匿名データ部会 議事概要

1 日 時 平成24年12月25日(火) 13:00~14:57

2 場 所 中央合同庁舎4号館4階 共用第2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、厚生労働省、農林水産省、経済産業省、国土交通省、東京都、千葉県

(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：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) 前回部会での質問に対する対応について

ア 匿名データの作成手順について

前回部会の「エ 情報の削除」の審議の際に、匿名データの作成手順について情報の削除とスワッピングの順番に関する質問があったことに対し、諮問者から、まず発生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を行い、その後スワッピングを行ってから、母集団一意又は二意のレコードを含む世帯の削除を行うとの説明があった。

イ 年齢差の大きい夫婦、年齢差の大きい又は小さい親子のいる世帯の削除の状況について

前回部会において、有用性の観点から「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」の削除の状況を確認したいという意見があったことに対して、諮問者から、0.5%基準による方法と今回のクロス表の分布をみて削除を行う方法とで、それぞれどの程度のレコードが削除されるかについて説明が行われた。各委員等の主な意見等は次のとおり。

- ・ 0.5%基準で削除した場合は、削除数が多過ぎてこの変数自身が使えなくなってしまう。ただ、今回の方法でも、年齢差が45歳以上の親と長子、親と末子のいる世帯の削除数は、まだ相当な数になるという印象がある。再婚や晩婚化の影響などによると思うが、女親の場合はあまり問題にならないものの、とくに男親の場合がまだ多く、心配である。
- ・ 年齢差が45歳以上の親と子は、地域ごとに見ると特定されやすくなっている可能性もあるので、外観識別性の観点から何らかのクロスを取ったときに危険性があるのかどうかを確認したい。
- ・ 年齢差の大きい夫婦、親子等のいる世帯については、地域差が非常に大きい、公表されている集計表からどの程度わかるのか。また、匿名データを提供する際に、どの程度削除されているかを、ユーザがどのくらいわかるのか。

→ 公表されている結果表から夫婦の各歳別の年齢差を算出することは可能だが、親と長子や親と末

子の各歳別の年齢差については算出することはできない。また、ユーザに対して、個々の方法によるレコードの削除数を公表することは想定していない。

- ・ 例えば、年齢差が 45 歳以上の親と子のいる世帯は地域差が非常に大きい、全体のサンプリングの中で、結果として地域によってはアンダー・サンプリングになるということがあり得るのか。
- 本来、削除すべきデータと考えているので、地域によってそのようなことが起こるのはやむを得ないと考える。
- ・ 年齢差の大きい親子のいる世帯については、子供が小・中学生の場合は学年がわかるので問題になるが、子供の年齢が上がればほとんどわからない。また、年齢差の大きい親子等のいる世帯の削除よりも、1%の抽出でサンプルから落ちる世帯の方が多いので、この部分を細かく議論しても結論は出ないのではないか。

#### 《部会長のまとめ》

- ・ 年齢差が 45 歳以上の親と子のいる世帯について、地域を限定した形でよいので、削除に関する資料を準備いただきたい。また、このような世帯は今後増えてくる可能性が高いので、匿名化の方法について今後も検討していただくことが必要ではないか。

#### (2) 前回審議できなかった残りの論点について

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 世帯員に関する項目の再編等について

##### (ア) 産業（大分類）

- ・ 「産業」の中では、「農業」、「林業」、「漁業」を統合することになっているが、「職業」の方では、「農林漁業作業員」を他の「保安職業従事者」や「運輸・通信従事者」と統合することが必要という計画となっている。「産業」と「職業」で農林漁業の扱いが異なるが、匿名性の観点から問題は生じていないのか。

→ 「産業」の中の「農業」、「林業」、「漁業」の統合については、地域別に統合後の農林漁業の就業者割合を確認したところ、一部の地域において割合が低くなることがわかった。このため、匿名化を担保するために何らかの措置が必要となることから、「農業」、「林業」「漁業」を他の「鉱業」及び「電気・ガス・熱供給・水道業」と統合するか、あるいは、一部の匿名性を確保できない地域の農林漁業を対象に、外観識別性の高い事項との間でクロス表を作成し、その分布に応じて特定化の危険性の高いレコードを含む世帯を削除する方法の2案を提案したい。

- ・ 農林漁業を他の産業と統合して「その他」のような形でまとめることは、産業の特性を考えると現実的ではない。農林漁業は統合するが、一部の特異なレコードを削除する案の方がよいと考える。
- ・ 農林漁業を統合することについては異論はないが、一部の地域で農林漁業の就業者割合が低くなっているため、このレコードを削除するということか。

→ 全てを削除するわけではなく、外観識別性が高いようなデータを洗い出して、削除できればと考えている。

- ・ 外観識別性が高い属性は何かを検討して、その上でクロス集計をして、できるだけレコードの削除数を少なくするようにしたいということかと思う。今回の計画には、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が含まれる世帯を削除することになっているが、その世帯が農林漁業の中の特異なレコードと重複している可能性もあり、思ったよりも削除の対象となる数は少ないのではないか。
- ・ 標準産業分類という視点から考えれば、第1次産業、第2次産業、第3次産業というのは一つの

意味を持っていると思う。ユーザサイドから見て、第2次産業、第3次産業と分けられないようなデメリットがあっても、原案の統合方法がよいかは疑問がある。「鉱業」と「建設業」の統合はかまわないが、第1次、第2次、第3次産業に区分できるように、「電気・ガス・熱供給・水道業」を「製造業」ではなく、例えば「情報通信業」と統合してはどうか。1次、2次、3次の産業区分で分けた方が、有用性が高いのではないか。

→ 各産業の職業大分類別の構成をみると、「情報通信業」は「専門的・技術的職業従事者」が非常に多く、「生産工程・労務作業員」が少ない。一方、「電気・ガス・熱供給・水道業」では「専門的・技術的職業従事者」は少なく、「生産工程・労務作業員」が多くなっており、これは「製造業」と似ている。また、従業者規模から見て、「電気・ガス・熱供給・水道業」は「製造業」に統合した方が妥当ではないかと考えたところである。

- ・ 産業分類の統合において、「その他」の区分を設けるという案も出ているが、将来的に国勢調査の匿名データを新たに作成する際に、その区分が匿名データにおける「産業」の分類に影響を及ぼす可能性があることから、「その他」の区分を設けない方がよいと思う。産業や職業に関しては、今回提案されているように、他の分類区分と統合するという方向性しかないのではないか。
- ・ 「鉱業」の統合先は、第2次産業ということもあり、やはり「建設業」ではないか。「電気・ガス・熱供給・水道業」は非常に割合が少ない一方、「製造業」に関連があるので、原案のとおりこれらを統合することでよいのではないか。
- ・ 「製造業」が全体に占める割合は地域によってかなり差があり、関心も高いので、これを他の分類と統合しない方がよいのではないか。「建設業」も同様だが、こちらは仕方がないと思っている。
- ・ 基本的にいくつかの分類を統合して「その他」の区分を設けるのは問題がある。「電気・ガス・熱供給・水道業」の統合を考える場合に、「製造業」の区分を単体で残しておくかどうかということの問題であると思う。
- ・ 今回の匿名データの作成では地域情報を提供することを強調しているので、そうであれば製造業を分析したいと考えるのではないか。
- ・ 今回は地域分析をしたいというユーザにも応えるように、「都道府県」及び「人口50万以上の市区」の地域情報を提供することになっているが、例えば諸外国でも地域区分の詳細なマイクロデータの提供において「産業」が提供されない場合がある。このため、今回「産業」をある程度統合することはやむを得ないと考えており、実際の産業の特性に近いと思われる区分で考えると、原案でよいのではないか。

#### ＜部会長のまとめ＞

- ・ 農林漁業を他の区分と統合すると著しく有用性を損なうため、割合が低くなる地域が限定的であるという前提の下に、外観識別性が高いと考えられるレコードを削除する案を採ることとしたい。
- ・ 「鉱業」及び「建設業」を統合することに関しては、適当と判断する。
- ・ 「電気・ガス・熱供給・水道業」の統合については、非常に難しい判断であり、「製造業」の区分を単体で残すことに対する配慮の必要性については課題として残るものの、原案のとおり、「製造業」と統合することとしたい。
- ・ また、「複合サービス業」と「サービス業（他に分類されないもの）」を統合することに関しては、適当と判断する。

#### （イ）職業（大分類）

- ・ 「職業」に関しては、「農林漁業関係職業」、「生産・運輸関係職業」、「販売・サービス関係職業」、「事務・技術・管理関係職業」の4部門に集約する場合があるため、統合に際してはこれに倣い、「農林漁業作業員」を別掲とし、「運輸・通信従事者」を「生産工程・労務作業員」に、また、

「保安職業従事者」を「サービス職業従事者」に統合することはできないか。

- 「保安職業従事者」、「農林漁業作業者」、「運輸・通信従事者」は、地域で更にクロス集計を行った場合に母集団一意が散見されること、「保安職業従事者」などは外観識別性も高いことから、統合が必要である。本来であれば、御指摘のように4部門に集約が可能なようにまとめられればよいが、この3つはいずれにせよ統合が必要なので、これだけで統合した方が、統合先は少なく、他の区分への影響を避けるという観点からはよいのではないかと考え、原案の方法でやむを得ないのでと考えている。

《部会長のまとめ》

- ・ 「保安職業従事者」、「農林漁業作業者」及び「運輸・通信従事者」を統合することについては、匿名化の観点からやむを得ない措置であり、認めることとしたい。

(ウ) 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」を統合することは、適当と判断する。

(エ) 利用交通手段（平成12年国勢調査）

- ・ 「勤め先・学校のバス」、「ハイヤー・タクシー」、「オートバイ」は、全く違うものなので、これらをまとめて更に「その他」と統合してしまってもよいのか。バス、車、二輪などの利用交通手段の形態により分けた方がよいのではないかと。
- 「勤め先・学校のバス」、「ハイヤー・タクシー」、「オートバイ」は、0.5%基準で見ても非常に少なく、外観識別性も高いので何らかの統合が必要である。統合に際してはなるべく類似性のあるものを統合した方がよいという観点から、利用交通手段別に産業大分類別の就業者割合を見てみたが、「オートバイ」と「自転車」は似ているものの、「乗合バス」と「勤め先・学校のバス」、「自家用車」と「ハイヤー・タクシー」はそれぞれ少し性格が違っている。これは職業大分類別に見ても同様であり、バス、車には必ずしも類似性があるとは言えないと考え、原案のとおりではないかと考えている。
- ・ 「勤め先・学校のバス」、「ハイヤー・タクシー」は非常に少ないので、「その他」とまとめてもよいかもしれないが、「オートバイ」と「自転車」は二輪で性格が似ていると考えられるので、これらを統合することはできないか。
- 「オートバイ」、「自転車」の統合及び残された「勤め先・学校のバス」、「ハイヤー・タクシー」、「その他」の統合については、それぞれ問題がないかを精査したい。

《部会長のまとめ》

- ・ 「オートバイ」と「自転車」の統合、及び「勤め先・学校のバス」、「ハイヤー・タクシー」、「その他」の統合について、次回資料を提示いただき、検討することとしたい。

イ 世帯に関する項目の再編等について

(ア) 世帯の種類、世帯人員、及び世帯の家族類型

- ・ 世帯の種類としては、「一般世帯」及び「施設等の世帯」の2区分を提供することとし、「施設等の世帯」の内訳は提供しないこと、「施設等の世帯」の世帯人員については提供しないこと、及び「世帯の家族類型」を6区分で提供することは、適当と判断する。

(イ) 家計の収入の種類（平成12年国勢調査）

- ・ 「家計の収入の種類」については提供しないという計画であるが、本部会や統計委員会でも提供できないかとする意見があったところであり、提供の可能性について検討していただきたい。



→ 「家計の収入の種類」は、忌避感の非常に強い調査項目であることから、調査本体に与える影響を懸念しているところであるが、既に平成 22 年国勢調査の調査項目からは除かれていること、匿名化措置を徹底して情報の漏えいを防ぐということが前提であるならば、提供は可能であろうと考えて、①「賃金・給料が主な世帯」の内訳の、「農業収入もある世帯」と「その他」を統合する、②「農業収入が主な世帯」と「農業収入以外の事業収入が主な世帯」を統合し「事業収入が主な世帯」とするが、それぞれの内訳は提供しない、③「内職収入が主な世帯」と「その他の収入が主な世帯」を統合し「その他（内職収入が主な世帯を含む）」とするが、その内訳は提供しない、④「恩給・年金が主な世帯」は統合せず内訳を含めそのまま提供するという再編案を準備した。調査本体に影響を与えないためにも、きちんとした統合が必要と考えている。

《部会長のまとめ》

- ・ 「家計の収入の種類」については、有用性の観点からは提供される方がよく、匿名性の観点からは、提案のあった分類区分の再編案による母集団に対する割合に問題は見られないことから、分類区分を統合した上で提供することが適当と判断する。

(ウ) 住居の種類・住宅の所有の関係

- ・ 「住宅に住む一般世帯」について、その内訳の「公営の借家」及び「都市機構・公社の借家」を統合すること、「給与住宅」及び「間借り」を統合すること、並びに「住宅以外に住む一般世帯」の内訳（「会社等の独身寮」及び「その他」）は提供しないことは、適当と判断する。

(エ) 住宅の床面積、住宅の建て方、建物全体の階数、及び世帯が住んでいる階

- ・ 住宅の床面積については、実数ではなく、基本的に既存の統計表に合わせた階級で提供することとし、この際、「200～249 ㎡」及び「250 ㎡以上」を統合すること、住宅の建て方については、「長屋建」及び「その他」を統合すること、建物全体の階数については、実数ではなく、基本的に既存の統計表に合わせた階級で提供することとし、地域区分ごとの出現頻度により、必要に応じて、「6～10 階建」、「11～14 階建」、「15 階建以上」の区分のうち、「6～10 階建」以上、あるいは「11～14 階建」以上の区分を統合すること、及び世帯が住んでいる階については、実数ではなく、基本的に既存の統計表に合わせた階級で提供することとし、地域区分ごとの出現頻度により、必要に応じて、「3～5 階」、「6～10 階」、「11～14 階」、「15 階以上」の区分のうち、「3～5 階」以上、「6～10 階」以上、あるいは「11～14 階」以上の区分を統合することについては、適当と判断する。

ウ 匿名化措置を予定していない事項について

- ・ 匿名化措置が予定されていない事項については、特に問題はなく、適当と判断する。

エ 他の情報との関係について

- ・ 今回の計画では、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等により、外部の情報との対応関係から世帯・個人を特定される危険性を可能な限り低くしていると考えられることから、適当と判断する。

オ その他

(ア) トップコーディングが行われた変数の基本統計量の提供について

- ・ トップコーディングが行われた変数の基本統計量については、就業時間について提供することだが、これは地域ごとに提供するのか。

→ 今後検討する予定である。

- ・ 地域別で提供することを検討していただきたいと思うが、その可能性を含めて、次回説明いただきたい。

(イ) 匿名データの提供時期の短縮化について

- ・ 調査実施後5年以上経過したものから提供することについては、異論はない。国勢調査は5年ごとに実施されるものであり、この提供時期を短縮化することは難しいと考えられる。特に、今回は地域情報を提供するため、秘匿性の確保は慎重にすべきであり、少なくとも5年は待つべきであろうと考える。5年周期の調査ではなく、むしろ3年周期や月次などの調査に対する課題ではないか。
- ・ ユーザがどのように捉えるかは様々であり、古いデータでは利用されないこともある。日本の統計の活用という問題で言えば、調査実施後5年以上経過したものでないと提供できないということであるならば、例えば、より古いデータについて提供するという考え方もあるのではないか。非常に古いデータまで匿名化されれば、それはそれで大変価値があると思う。

《部会長のまとめ》

- ・ 一般論として、匿名データの提供時期は早期化するのが望ましいが、国勢調査については、悉皆調査であることから、調査客体の信頼性を確保し、匿名性を確保する観点からも、提供までに十分な期間を置いた方がよいと考えられるので、調査実施後5年以上経過したものから提供することは適当と判断する。

(3) その他

- ・ 次回の匿名データ部会は、1月31日(木)10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スワッピングに関するユーザへの説明の仕方については、次回議論することとされた。

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＞