

第 25 回 匿名データ部会 議事概要

1 日 時 平成29年12月25日（月） 14:00～15:20

2 場 所 総務省第二庁舎 6階 特別会議室

3 出席者

（部 会 長） 北村 行伸

（委 員） 川崎 茂、永瀬 伸子

（専 門 委 員） 神林 龍、寺村絵里子、南 和宏

（審議協力者） 総務省（政策統括官（統計基準担当））、経済産業省、国土交通省、
千葉県

（諮 問 者） 厚生労働省：政策統括官（統計・情報政策担当）付参事官付審査解
析室：田中室長ほか

（事 務 局） 総務省：横山大臣官房審議官
総務省統計委員会担当室：上田次長ほか

4 議 事

（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

冒頭、厚生労働省から匿名データ作成に係るリサンプリング方法等について再説明があった。

その後、答申案について審議が行われ、一部文言等の修正を行うことを前提に了承された。答申案については、所要の修正を行った後、第 118 回統計委員会（平成 30 年 1 月 18 日開催予定）において部会長から報告することとされた。

リサンプリング方法等に係る主な質疑は以下のとおり。

○資料では、本体調査が採用している集落抽出のメリットばかりを記述しているが、デメリットもある。二段抽出方法等と比べるとコストパフォーマンスなどは劣る可能性が高い。また、「出現頻度の低い母子世帯等を漏れなく把握可能」という記述は、言い過ぎではないか。

○「調査区から集落が特定できる」、「集落が判明しても」との記述があるが、地域情報を提供していないので、この懸念はなく、このようなことを想定する必要はない。

○リサンプリングの第 2 段階である調査区から一律 8 割の世帯を抽出することは、匿名化措置として意味がない。調査区を抽出していることで匿名化は担保されており、さらに 8 割を抽出する必要はない。

- 仮に、隣の世帯の属性が分かっていたとしても、地域情報を削除しており、同じような属性を持った世帯は、全国の他の地域にも存在するかもしれないので、匿名性は確保されていると理解している。
- 都道府県によっては、調査されている調査区の地点のうちごく一部の調査区の抽出により匿名データが作成されている。この方法では集落の特性の偏りが反映されやすい。調査区単位で抽出するよりも直接、世帯を抽出した方が世帯属性の分布などが偏る危険性が少なくなる。また匿名性の確保も高まり地域表章の可能性も高まるのではないか。
 - 今後、ダイレクト・リサンプリングの方法についても検討する。
- 世帯員数8人以上の世帯をトップコーディングしているが、どのような基準で人数を決めているのか。例えば、5人以上でトップコーディングすれば、地域情報を提供できるのではないか。
 - 今後、地域情報を付与したデータを検討する際に勉強してみたい。
- リサンプリングを見直すと、これまでの作成方法とは違うデータになる。時系列でみると、同じ手法ではなくなるので、変更した箇所の情報を示していただきたい。
 - リサンプリングについては、大まかな情報を出していきたい。
- ホームページ上にあるリサンプリングの説明文は、5～6行の記述しかなく十分な情報とはいえない。
 - ホームページ上は公開情報として提供できる範囲とし、匿名データの利用者には、より詳細な情報提供をするようなことを考えたい。
- 作成方法が大幅に変更になったときは、データ遡及も含めて考えるべきではないか。
 - マンパワーなどを考えると、いまは難しい。
- 時系列の問題は悩ましい。匿名データについては、ベストな作成方法というものはないのではないか。作成方法を変えることに躊躇してはいけない。ベストな方法に向けて変えていくスタンスを持つべきである。
- 匿名データを利用する際には、本体調査の特徴を理解しておかないとミスリードする。平成27年の国勢調査結果では、世帯数は増加が続いているが、国民生活基礎調査では、減少となっている。このようなことを理解して利用する必要がある。
- 国勢調査とのかい離の原因は、何か。
 - 若年層や単独世帯においてかい離があると認識している。
- 匿名データを作成するに当たり、本体調査の分布に合わせるとしているのに、本体調査に歪みがあるのは問題。本体調査ではなく、より実態に近いものをターゲットにすることもあり得るのではないか。
 - 検討してみたいが、難しい問題である。

(部会長のまとめ)

- 本体調査の標本設計が、匿名データのリサンプリングに制約をかけていること

が分かった。本部会では、本体調査には言及できないが、統計委員会において、その旨、報告したい。

- リサンプリングについては、ダイレクト・リサンプリングの方法に見直すべきなど課題があるので検討すべき。
- ただし、リサンプリングの見直しには時間も必要で、リソースの制約も考慮すると、今回、従来どおりの方法でリサンプリングすることはやむを得ない。
- 今後、リサンプリングを見直す際には、匿名性の確保、地域情報の付与、過去への遡及、時系列の整合性、匿名化手法の情報提供の在り方などについて検討すべきである。
- なお、次回に向けた検討については、一定程度方向性が定まった段階で、匿名データ部会に何らかの形で報告し、意見交換させてほしい。

答申案に係る主な意見は以下のとおり。

- ダイレクト・リサンプリングの優位性として、「作業効率の観点」が上げられているが、むしろ「統計精度の観点」ではないか。
- 「世帯員単位での抽出」とあるが、単に世帯員単位で抽出するというだけでなく、世帯員単位での所得情報の提供可能性にも検討すべき。
- 地域情報の付与を検討する際には、外観識別情報について再整理することが必要である。
- 今後の課題として、ダイレクト・リサンプリングの方法について検討するとあるが、それも含めた匿名化手法をバランスよく見直すということだと思う。その意味では、統計精度向上、地域情報の提供、世帯員単位での抽出などの観点から、リサンプリングを含めて検討するという趣旨にすべき。
- 国民生活基礎調査に関する既存の統計委員会答申（平成28年1月21日答申）で指摘されている課題にも留意して検討することを記述すべきである。
- 「今後の課題」は、「地域情報の付与及び再抽出の単位」だけに限ったものではなく、もっと広範囲の観点で記述すべきである。
- 8割で抽出していることも含め、もっとフリーに検討するニュアンスを表現できればよいと考える。

以上

<文責 総務省統計委員会担当室 速報のため事後修正の可能性あり>