

第 23 回 匿名データ部会 議事概要

1 日 時 平成29年11月14日（火） 9:30～10:50

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） 寺村絵里子

（審議協力者） 総務省（政策統括官（統計基準担当））、総務省統計局、財務省、
文部科学省、経済産業省、国土交通省、東京都、千葉県

（諮 問 者） 総務省：統計局統計調査部調査企画課 栗田課長ほか

（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）住宅・土地統計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）住宅・土地統計調査に係る匿名データの作成について

総務省統計局から諮問の概要について説明の後、論点メモに沿って審議が行われ、概ね適当と判断された。今回は、匿名データを含めた調査票情報の提供について、基本計画での検討状況など事務局からの報告及び答申案を審議することとされた。

主な意見は以下のとおり。

① 匿名データの利用状況、役割

○作成済みの匿名データ及び法第33条の規定に基づく調査票情報の提供状況を教えてほしい。

→本調査の匿名データの提供状況は、提供開始年の平成22年度は6件で、その後は毎年数件程度で推移、33条での提供は、年間数十件で、多くは地方公共団体への提供である。匿名データの利用用途は、高齢化社会や住宅問題の研究分析となっている。

○匿名データ作成は、手間隙をかけているのに利用が少ない。どうして少ないのか、原因が何か、考える必要があるのではないか。

○匿名データは、サンプル数も多いのがメリットである。もっと利用されてもいいのではないかと思う。

- 匿名データの作成は、委員の御示唆も踏まえ利用者の利便性の向上も図りながら、取り組んできたところ。匿名データが多くの方に利活用されるよう、今後も引き続き利便性の充実や早期提供に努力していきたい。
- 匿名データ作成方法としては、地域間でデータを入れ替えるスワッピング手法などを取り入れると、もっと小さなサンプルでも作成が可能になるのではないか。
 - 現在の作成方法は、匿名性の確保に重点をおいており、そのために詳細な地域情報などがなく使い勝手がよくない。オーダーメイドに近い匿名データを作成すればいいが、そのためにはリソースも必要で大変である。
 - 今回の住調匿名データの議論はともかく、そもそも匿名データは、統計改革でどのように制度が見直されるのか。
- 匿名データの提供先については、従来の学術研究等に資する統計の作成等に加え、一定の公益性を要件に民間企業にも提供できないか法改正を検討している。そのためには、個人情報保護や調査現場とのバランスの問題もあると考えている。
- 誰にでもオープンにする必要はないが、使いやすく、up-to-dateなデータが提供できるといいのではないか。データへのアクセスをよくすることが大事である。
 - 匿名データのターゲットをどのような人にするのか検討すべきである。

(部会長のまとめ)

- 匿名データを含めた調査票情報の提供について、基本計画での検討状況などを事務局で整理して、次回、報告してほしい。

② 都道府県別トップ（ボトム）コーディング

《諮問者の説明》

データの有用性を高めるため、トップコーディング、ボトムコーディングの基準値を全国一律から都道府県別に設ける。

主な意見は以下のとおり。

- トップコーディング、ボトムコーディングを都道府県別に設定することはいいことである。既に提供している年次にも適用できないか検討してほしい。
- 都道府県別の設定は、家賃などは明らかに県によって異なるので、よい措置だと思う。
- 都道府県によって、建物面積や家賃などの分布が異なるので、この措置で有用性が高まると思う。
- 都道府県別に設定することは当然かと思う。サンプル数も多く有用な情報を提供できるものとする。

(部会長のまとめ)

- 都道府県別のトップコーディング、ボトムコーディングの基準値の設定は、適当と判断する。

③ 東日本大震災に関連する項目

《諮問者の説明》

平成25年新規項目のうち「東日本大震災による転居」及び「東日本大震災による改修工事」の回答状況をみると、分布は一部の地域に限られるが、それらの世帯において複数の事項と組み合わせることにより、世帯が特定されるリスクが高まることが考えられるため、情報提供は行わないことにしたい。

主な意見は以下のとおり。

- 東日本大震災の影響は注目される論点だが、提供できない理由は理解した。
- 東日本大震災でどれだけ移動があったのかは、大きな関心事であるが、33条での利用を考えると、匿名データでそこまでの情報提供は必要ないと考える。

(部会長のまとめ)

- 平成25年新規項目のうち「東日本大震災による転居」及び「東日本大震災による改修工事」の情報を提供しないことは、匿名性の確保の観点から適当と判断する。

④ 複数の匿名データの作成

《諮問者の説明》

地域を詳細化した複数の匿名データ作成には、匿名レベルの考え方や秘匿方法など基礎的な検討課題が多い、現在、平成22年国勢調査の匿名データ作成の検討の中で、並行的に分析を進めており、その検討結果も踏まえた上で、次回検討したい。

主な意見は以下のとおり。

- 平成22国勢調査の匿名データ作成の検討について、今後のスケジュールを教えてください。

(部会長のまとめ)

- 現段階では、複数のデータを作成できないことは適当と判断するが、国勢調査の匿名データ作成に関する今後の見通しを次回示してほしい。

⑤ 匿名データの提供時期の短縮化

《諮問者の説明》

調査実施後5年以上経過したものを提供していたが、平成20年のデータに関しては遅かったと認識している。今後、提供時期までの期間の短縮について検討したい。

主な意見は以下のとおり。

- 可能な限り早く提供してほしい。調査結果公表後、速やかに提供できるような形

になるのがいいのではないか。

(部会長のまとめ)

○今回の措置は適当と判断するが、速やかな提供に努めること。

⑥ トップコーディング等が行われた変数

《諮問者の説明》

多変量解析等に資するため、トップコーディング等が行われた変数については、その統合された部分に関する平均値と標準偏差を都道府県別に提供する。

主な意見は以下のとおり。

○平均値と標準偏差だけでいいのかと思うが、大丈夫か。

→昨年諮問審議された就業構造基本調査でも年齢の部分を同じように平均値と標準偏差を提供している。

(部会長のまとめ)

○今回、トップコーディングした変数の平均値と標準偏差を提供することは適当と判断する。ただし、今後、トップコーディングした変数についての情報提供については、匿名データ全般の課題として、機会を改めて検討したいと考える。

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>

第 24 回 匿名データ部会 議事概要

- 1 日 時 平成29年11月29日（水） 17:00～19:00
- 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））、総務省統計局、財務省、文部科学省、厚生労働省、経済産業省、国土交通省、東京都、千葉県
 - （諮 問 者） 総務省：統計局統計調査部調査企画課 栗田課長ほか
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）住宅・土地統計調査に係る匿名データの作成について
 - （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

（1）住宅・土地統計調査に係る匿名データの作成について

冒頭、現在検討されている次期基本計画における匿名データを含めた調査票情報の提供の検討状況について、事務局と総務省統計局から報告があった。

また、総務省統計局から、複数の匿名データを作成することについて、次期基本計画を先取りして、統計研究研修所の支援を受けて、具体的な検討を開始している旨、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 117 回統計委員会（平成 29 年12月 19 日開催予定）において部会長から報告することとされた。

答申案に係る主な意見は以下のとおり。

- 新規追加項目に関する最終行の「提供しないことは適当である」の記述は、ネガティブな印象を与えるので、何か表現を工夫すべきではないか。
- 検討したい。

○複数の匿名データ作成に関して、「匿名レベル」は、あまり聞きなれた言葉ではないので、表現を変えるようなことが必要ではないか。また、「国勢調査の匿名データを利用した」とあるが、国勢調査の匿名データを利用して検討するのではなく、国勢調査の匿名データを作成するための検討プロセスの中で得た情報などを利用して検討するということだとすれば、言葉が足りない気がする。

→検討したい。

○複数の匿名データ作成に関して、「より詳細な地域情報」とあるが、都道府県レベルより詳細などと比較するレベルを書く必要がある。

→検討したい。

(2) 国民生活基礎調査に係る匿名データの作成について

厚生労働省から諮問の概要について説明の後、論点メモに沿って審議が行われ、次回、引き続き審議することとされた。

主な意見は以下のとおり。

○リサンプリング方法については、見直してもいい時期にきているのではないか。ある世帯が調査対象になったら、その近隣世帯の情報が分かりやすくなるという説明は、根拠が薄弱である。地域情報を提供しないことを前提にすれば、ダイレクトサンプリングでも構わないと思う。今回の作業で間に合わなければ、次回までに検討してほしい。

○抽出された集落は悉皆だというのが、リサンプリングされた結果、全体の2割しか抽出されていないのであれば、全国全体から抽出してもよいと思う。

○今後、リサンプリングを見直すのであれば、抽出率の8割や標本一意であったときに母集団でどのくらいの確率で一意になっているのかを評価するとよいのではないか。

○匿名データCを作成して他の匿名データとマッチングすると、かなりのデータが特定される可能性があるとの説明だが、同じ世帯が選ばれているということが認識できる可能性はあるが、どこの世帯かという秘匿性は破られてはいないので、その説明は誤解を招きやすい。

○本体調査が二段階抽出なので、匿名データも二段階抽出をしているが、効率性はよくないし、むしろ特定されるリスクが高いのではないかと考えるので、検討に時間を要するのであれば、引き続きの課題にしてもよいと思う。

○外観識別性の高い属性は何かを定義し、その属性で度数分布表を作成し、そこに地域情報を入れる、入れないで、一意識別がどう増えるかを検証すれば、秘匿性の確保が定量的に分析できると思う。

○次回に説明される際に、第一段階のリサンプリング率のマックス8割と第二段階の一律8割について、どうして固定なのか説明してほしい。(神林専門委員)

○都道府県よりも粗い地方ブロックのような地域情報は、医療を専門に研究される方にとっては使えない情報とのことだが、使えなくてもリサンプリングには役に立つかもしれない。

- 所得の内訳を提供していただくことは有難い。欲を言えば、世帯主と配偶者の内訳が分かると、もっとよい。
- 所得の内訳で、雇用者所得と公的年金・恩給以外の項目を提供しない理由は何か。また、世帯の総所得は提供せずに、世帯員の所得だけを提供するという選択肢はないのか。
- 世帯員の情報を提供することは、リサンプリングの方法に関連するが、世帯員でのリサンプリングは困難と考えている。

(部会長のまとめ)

- 新規に追加された事項を提供すること、所得の内訳情報として「雇用者所得」と「公的年金・恩給」を提供することは適当と考える。ただし、リサンプリングの方法とそれに関連して地域情報の付与については、次回、もう一度議論したい。

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>