

第6回匿名データ部会 議事概要

1 日 時 平成23年3月8日(火) 14:58~16:47

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、津谷典子委員、廣松毅委員、伊藤伸介専門委員、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、日本銀行、東京都、千葉県

【諮問者(厚生労働省統計情報部)】

中島企画課審査解析室長、山田企画課審査解析室室長補佐、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) 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

(1) 第43回統計委員会における統計委員会委員の意見について

事務局から、資料2「第43回統計委員会における統計委員会委員の意見」が紹介された。

(2) 前回部会での指摘を踏まえた秘匿措置の見直しについて

諮問者から、前回の部会審議で指摘された事項に対して、検討の結果、秘匿措置を見直したい旨の回答があり、委員から当該回答に対する意見等が述べられた。

各委員等の主な意見等は次のとおり。

厚生労働省の回答

- ・ 裾切りによるレコードの削除のうち、「年齢差の大きい親子」の含まれる世帯に加えて「年齢差の小さい親子」の含まれる世帯も削除することとしたい、「同一年齢の子供が3人以上いる世帯」について、世帯員の年齢が年齢階級別に提供されることを踏まえ、「同一年齢階級に4人以上いる世帯」のレコードの削除へ変更することとしたい。
- ・ リコーディングを予定していた項目のうち、世帯票の「希望する仕事の形」、健康票の「悩みやストレスの原因」、「最も気になる悩みやストレスの原因(主要因)」及び「健診を受けなかった理由」について、専ら意識を問う項目であり、外観から識別される可能性が低いことから、そのまま提供することとしたい。

委員等の意見

- ・ 「年齢差の大きい親子」について、10階級以上離れている場合はレコードを削除することとしてい

るが、父親と子の年齢差に関しては10階級離れていてもそれほど稀な事例とは言い切れない。父親と母親のそれぞれで、子との年齢差を確認すべきではないか。

部会長のまとめ

- ・ 「年齢差の小さい親子」の世帯のレコードを削除の対象とすることは適当。なお、「年齢差の大きいまたは小さい親子」の世帯については、父親や母親で層別に見て0.5%を超えるような時に秘匿措置の緩和が可能であるか、実施者には確認していただき、結果については部会長一任としたい。
- ・ 「同一年齢の子供が3人以上いる世帯」について、「同一年齢階級に4人以上いる世帯」の削除に変更することは適当。
- ・ 「希望する仕事の形」以外に、「悩みやストレスの原因」、「最も気になる悩みやストレスの原因（主要因）」及び「健診を受けなかった理由」に係る秘匿措置を緩和することは、有用性の観点から適当。

(3) 答申(案)の審議について

部会長から、資料1「諮問第34号の答申『国民生活基礎調査に係る匿名データの作成について(案)』」が示され、事務局が答申(案)について項目ごとに朗読の後、審議が行われた結果、一部で所要の修正を行うこととされたが、答申案は概ね適当であるとして採択された。

なお、答申(案)の修正文の表現については部会長に一任することとされた。
各委員等の主な意見等は次のとおり。

ア 裾切りによるレコード削除(2の(1)のウ)

委員等の意見

- ・ 当初の計画になかった「年齢差の大きいまたは小さい親子」について、部会審議では「年齢差の大きい親子」を提供することは問題がないかとの指摘があったが、実施者による検討の結果、「年齢差の小さい親子」についても問題が見つかったので、こういった部会の審議経過を正確に記述する意味で、案2が適当ではないか。

部会長のまとめ

- ・ 「年齢差の大きいまたは小さい親子」については、案2を採用することとしたい。
- ・ 同一年齢階級に一定以上の者がいる世帯について、レコードを削除することについては、「一定以上」ではなく「4人以上」と明確にすべき。

イ 出現頻度の低い選択肢のある項目(2の(2)のイの(イ))

委員等の意見

- ・ 「専ら本人の意識を問う項目であって」の部分に対し、削除の意見と「本人の意識に関わる項目でもあって」という修正意見の2つの意見があるが、この部分は本人の意識上でしかわからないから緩和できるという意図があり、単純に外観識別可能性が低いというだけでは問題。
- ・ 「意識を問う」となると一種のアンケート調査、意識調査というような位置づけになるように受け取られる可能性があるので「意識に関わる」としてはどうか。
- ・ 「本人の意識に関わる」とすると、他にも意識に関わる項目があるが、それは緩和できないので、「意識を問う」としたものである。

部会長のまとめ

- ・ 「専ら本人の意識を問う項目であって」については、原案通りとすることとしたい。

ウ トップコーディング等が行われた変数（3の（5））

厚生労働省の回答

- ・ トップコーディング等が行われた変数の平均値等の提供については、まだ提供実績もなく統一的・基本的考え方も整理されていないこと、該当世帯数が少ない場合に秘匿措置として十分かどうかやサンプルによってバラツキが非常に大きいなどの検討すべき事項があることが判明したことから、直ちには提供できず、検討の時間が必要。

総務省の意見

- ・ 総務省の4調査について、技術的にはどのくらい詳細な区分で出すのか、また、該当する対象が少ない場合は秘匿が破られる危険性があるという問題があるが、ご要望のあったデータを出せるよう検討したい。

委員等の意見

- ・ 公表統計と違った結果が出たときに、それが論文としてアクセプトされるのかという心配がある。今後、分野によってはあるかもしれないが、総務省4調査のときには当該平均値等がないから論文がアクセプトされなかったということは聞いたことはない。本来はオーダーメード集計で集計すべき情報ではないか。
- ・ トップコーディング等が行われた変数の平均値等については、メタデータという形で提供する方法もあるが、対象サンプルのレコードの中にトップ（ボトム）コーディングの閾値ではなく、当該部分の平均値等を入れるという方法も考えられるので、海外の提供事例も含めて検討する必要がある。
- ・ オーダーメード集計でできる場合もあるだろうが、匿名データとして利用する際に、トップコーディングした部分をインピュートしようとしてもどのような値を入れればよいのかわからないので使い勝手が悪いという意味であるならば、現時点では結論は出せず、更に検討が必要。
- ・ 対象の数がどの程度であればよいか、どのような情報の開示の仕方がよいのかについて技術的検討が必要である。ただ、匿名データの有用性の観点からすると、開示のリスクがそれほどないのであれば、その部分の補助情報を提供することは有用性があると思う。
- ・ トップコーディングした部分の平均値はかなり不安定になっている。サブサンプル数だけでなく、リサンプリングによるところも大きい。このまま公表されるのはむしろ問題であり、ミスリードされ、公表統計と違った結論を出されることを危惧する。
- ・ トップコーディングした部分の平均値を出すことは極めてミスリードを起こす可能性があることに注意した方がよいが、トップコーディングした変数全体の基本統計量であれば出せるのではないか。総務省4調査を含めて、統一的な基本統計量を出すことについての情報開示の方法論を検討していただきたい。

部会長のまとめ

- ・ 「トップコーディング等を行った変数の平均値等の提供可能性を速やかに検討する」の「平均値」を「基本統計量」に変更する。
- ・ この点については、総務省4調査のときも今後の課題となっていたので統一的に検討していただきたい。また、今後、他の基幹統計調査において匿名データを作成する際にも同じ問題が発生するため、基本統計量を提供することについて恒常的に配慮していただく必要があるということ、次の統計委員会において答申案を説明する際に、資料2の回答として報告したい。

エ その他の意見について

委員等の意見

- ・ 「3 今後の課題」に関して、「(1) 地域区分及びリサンプリングの単位」及び「(2) 所得票の情報の提供」の中で、潜在的な利用者のニーズにも配慮されるように「利用者のニーズを十分に考慮したうえで」と修文した。また、「匿名データA及び匿名データBの閾値」については、今後の課題に書くには少し技術的すぎる指摘ということで削除することとしたい。
- ・ 所得等の内訳の提供についても、総務省4調査とも連携を図って検討していただきたい。
- ・ 「調査客体」という言葉が多いが、「調査対象」とか別の表現を使っている場合もある。他の答申でどのように表現しているか見て検討してほしい。

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>