

第20回 匿名データ部会 議事概要

- 1 日 時 平成28年3月25日（金） 10:00～11:45
- 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) 北村 行伸
(委 員) 川崎 茂
(専 門 委 員) 川口 大司、南 和宏、村田 磨理子
(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、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 植山 克郎課長ほか
(事 務 局) 内閣府統計委員会担当室：佐々木 健一企画官ほか
- 4 議 事
(1) 就業構造基本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
事務局が匿名データ部会の趣旨と非公表の扱いを説明した後、就業構造基本調査の匿名データ作成について、諮問者からの説明後、論点に従い審議した。委員等の主な意見は以下のとおり。
(1) 匿名性及び有用性の確保
①世帯員の年齢の匿名化措置
・年齢のトップコーディングを90歳以上とすると区分によっては人数が少なく危険とのことだが、85歳で切っても区分によっては人数が少なく危険になることはないのか。
→地域が3大都市圏か否かの2区分のみで提供されていることから、危険性は一定程度低減していると考える。
・トップコーディングは85歳以上でよいと思う。産業分類や職業分類はかなり詳細に分けているが、今後は、産業、職業分類をもっと大きくくりにして、他の属性を細かくしたファイルを提供することも検討すべきではないか。
→年齢を重視したり、産業を重視したり、ニーズが色々ある。次回の作成の際には御指摘の点についても検討したい。
②19年調査において変更された項目の匿名化措置（初職に就いた時期、年齢）
・初職に就いた年齢については、リスクがあるなら提供しなくてもいいと思うが、このような集計事項をどの程度加工し、提供するのがいいのか、今後検討してほしい。ユーザーにとっては、調査票に近い情報があれば、自分で何とか出来る。また、初職に

就いた時期は、可能なら、過去10年を超えた部分についても5年くりにすることなく年単位で提供してもいいのではないかと。

- ・調査票では初職に就いた時期は年月単位となっている。そのままでは問題があるのか。報告書の表章区分に合わせる理由を教えてください。
- 一般的に、初職に就く月は多くが4月で、ほかの月は珍しいため、初職に就いた月の情報提供は避けた方がよいと判断した。初職に就いた時期は、最近10年については各年で区切っているが、10年以上前はニーズが少ないことから、結果表章についても5年区分にしている。
- ・10年以前も各年で出せないか。
- 明確なリスクを確認したわけではないが、各年よりも5年区分にして幅を持たせるとリスクが減る。
- ・利用者からすれば、計算などは簡単な方がよい。あるところから5歳区切りになるのは使いにくい。年単位でも支障はないのではないかと。

【部会長のまとめ】

- ・匿名性と有用性の確保については、おおむね適当と判断したい。
- ・初職に就いた年齢の提供については、次回部会で再度検討したい。

(2) 前回答申の「今後の課題」への対応について

諮問第13号に対する答申における「今後の課題」への対応について、総務省統計局は、「(1) 複数の匿名データの作成」については、検証の結果、現行匿名データの一部から再抽出する方法では、現行ファイルと再抽出したファイルをマッチングすることで匿名化が破られるリスクが高まるので、引き続き別の方法による作成の可能性について検討する、「(2) 匿名データの提供年次の緩和」については、調査客体への配慮から直近の調査結果については匿名データを提供しない、「(3) トップコーディング等が行われた変数の平均値等の整備」については、年齢の平均値等を利用の手引きなどに掲載したい、との説明を行った。続いて、欠席した永瀬委員からの意見が事務局から紹介された。

これに対する質疑は以下のとおり。

- ・研究者を育成する観点から考えると、分析に有用な地域情報や年齢各歳別情報などを付加した小さなサイズのファイルを作成することも検討してほしい。
- ・海外の外国人にも利用し易いデータ提供を考えてほしい。
- ・複数の匿名データの作成について、マッチングのリスクが大幅に高まる項目を外して匿名データを作成してみてはどうか。また、年齢を各歳で提供していないのは、年齢構成で世帯が特定されるのを避けるためのはずなので、検証の仕方としては、世帯員の年齢構成のセットからユニークなものが導ける頻度でみるべき。
- ・複数の匿名データの作成について、ここまで安全性をみなくてもいいのではないかと。また、全部の属性を出すのではなく、特定の属性を削除したファイルを提供する方法もある。
- ・複数の匿名データの作成について、最初にリサンプリングした8割のレコードのみから抽出する方法に限界があるなら、それ以外から抽出を検討すべきではないかと。

→まったく提供されないデータの数が2割を下回ると、最初にリサンプリングする8割のデータのリスクが上がるため、最初のリサンプリング率を減らして、提供しないデータを2割確保する必要がある。調査年次の数回分を併せて利用するニーズが多いことがわかっているため、今回、そのような方法は採らなかった。

- ・リサンプリング率8割というのは、海外の事例と比べても高い。そこにこだわると、解決策がなくなる。

【部会長のまとめ】

- ・複数の匿名データの作成については、今回は見合わせるとのことだが、リサンプリング率8割の是非も含め複数ファイルの作成方法を引き続き検討していただきたい。
- ・調査実施後5年以上経過したものを提供するという基準については、やむを得ないと判断したい。
- ・トップコーディングした85歳以上の年齢層の平均値を提供することは適当と判断したい。
- ・なお、海外ユーザー向けの工夫や複数ファイルの作成を含め利用し易いファイル作成などの問題は、匿名データ全般の問題として、今後、検討していくことが必要である。

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>