

第19回 匿名データ部会 議事概要

- 1 日 時 平成27年1月8日(木) 10:00～11:10
- 2 場 所 中央合同庁舎4号館共用第3特別会議室
- 3 出席者
(部会長)北村 行伸
(委員)川崎 茂、津谷 典子
(専門委員)伊藤 伸介
(審議協力者)公益財団法人統計情報研究開発センター 村田 磨理子主任研究員、
総務省(政策統括官(統計基準担当))、財務省、文部科学省、農林水産省、
経済産業省、国土交通省、千葉県
(諮問者)厚生労働省大臣官房統計情報部企画課審査解析室：田中 伸彦室長ほか
(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) 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

前回部会で指摘された事項について厚生労働省から以下の回答があった。

「手助けや見守りの要する者の状況」における「主な介護者」のうち「その他の親族」と「その他」は、別々の区分として提供

「在学したことがない」は、「小学・中学」以下を「卒業」したと見做して提供
所得等の内訳情報は、今後提供する方向で検討

このうち については、在卒の学校の種類名称を「小学・中学以下」に変更することとされ、
その他の事項は、了承された。

答申案については、今回の部会審議における意見等を踏まえて修正することとされた。

委員等の主な意見は以下のとおり。

(1) 論点の審議について

「在学したことがない」の扱いについて

- ・匿名データ提供の手引き等に「『在学したことがない』と答えた人は一番低い学歴層である『小学校・中学校以下』に含めている」旨の説明があれば利用者の助けになる。

所得等の内訳情報の提供について

- ・所得の内訳の種類が一つしかない世帯がかなりの割合を占めていることが内訳情報提供

の制約になり、カテゴライズ化してもトップコーディング化しても匿名化を確保することは難しいと理解できた。所得の内訳にノイズを入れるといったパータベーションの手法が考えられるが、技術的には検討が必要ではないか。データの分布をさらに細かく見ながら検討を進めていく必要性を考えると、今後の課題として検討願いたい。

- ・内訳所得の中でも外観識別性の高いものと低いものがあるので、全ての内訳所得を一律に扱うのではなく、めりはりをつけて秘匿措置を行うことで情報の提供が可能になるのではないか。
- ・内訳所得と他の属性とのクロス集計が可能になることを考慮すると、地域の特定が出来ないとしても、匿名性の確保の検証については慎重に考える必要がある。今後の課題として明記すべきではないか。
- ・一つの調査の担当だけではなく、同様の問題を持っている担当が集まって技術的な問題を検討する場が必要ではないか。技術的なバリエーションや、調査によって適当な手法などの違いはあるか、どのような選択肢があるのかを検討することが必要だと考える。

提供時期について

- ・国民生活基礎調査がマスターサンプルとなって、その中から再抽出されて他の調査の対象となる世帯があるということは、調査が一回終わったとしても継続性があるので、識別性が上がるリスクが高い。このため、提供期間を5年程度開けることは、本件については特に必要だと考える。
- ・経年効果の観点からみると、3年経過した場合と5年経過した場合とで、秘匿性がどの程度変わるのかを検証するのは難しいが、集落抽出の影響を考えると、秘匿性保持の観点から3年よりも長い期間は必要だと考える。
- ・匿名データのユーザーの観点から見ると、最新の国民生活基礎調査は公表された集計表を見てからマイクロデータの分析を考える時間が必要なため、今のペースの提供で不都合はない。

【部会長のまとめ】

- ・提供時期については、昨年19年調査が提供され、今年、10年調査と22年調査が提供されるという状況なので、年次追加は着実に進んでいると評価し、適当と判断する。

(2) 答申案について

- ・2の(3)の「在学したことがない」について、「『在学したことがない』を選択した意思を尊重し」とあるが、回答者の意思については確認できないことや、匿名データの議論から外れることを考えると、この表現は避けた方が良い。
- ・「『在学したことがない』を選択した意思を尊重し」の代わりに「この情報の有用性に鑑み」などとし、「在学したことがない」を「卒業」に変更し、学校の種類の「小学・中学」の区分の名称を「小学・中学以下」に変更して措置すべき、とすれば比較的収まりが良くなるのではないか。また、「このことを利用者に向けてきちんと説明をする」趣旨の文言を加えても良いのではないか。
- ・(4)のアの「調査対象数、リサンプリング数が少ないことから今回は見送ることとして

いる」については、単純に「リサンプリング後の標本の規模が小さいことから」とした方が良いのではないか。

- ・(4)のアに「地域情報を付与した新たな匿名データ(Cデータ)」と記されているが、Cデータについては特段記載の必要はないのではないか。
- ・(4)のイの所得等の内訳情報について、第1段落の趣旨が、今回の審議に参加していない人には分かりづらい内容ではないか。第1段落では丁寧に検討したことは書いてあるが、何を検討したかは分かりづらい。経過を詳しく書くよりも、内訳項目の提供を検討したことを簡潔に記して次の段落の「しかしながら」につなげた方が良いのではないか。
- ・(4)のイの「一定の研究・検討」については、「慎重に行う」というニュアンスが不足しているのではないか。「一定の」を外し、「より慎重かつ精緻な研究・検討」とすると審議結果がより反映されるのではないか。
- ・(4)のウで作成対象年次の拡大については記述されているが、提供時期についても審議しているので、それも含めて記述すべき。
- ・3の(1)に「世帯員単位でリサンプリング」が出てくるが、これは重要課題と位置づけ過ぎない方が良いのではないか。世帯員単位に抽出することによって標本規模を大きくできるのであれば、また世帯単位の抽出では入手できない情報がある場合はメリットがあるが、同じ標本規模であれば有用性が低下するのではないか。
- ・その点については、地域情報の提供と世帯員単位のリサンプリングをセットにして考えるのであれば、「匿名データの利用者のニーズを踏まえつつリサンプリング等の匿名化措置を検討することによって何らかの地域情報を付与することの妥当性と可能性について検討する必要がある」などの形にして、地域情報の付与の検討と、世帯員単位のリサンプリングの方法の検討につなげることも考えられる。
- ・世帯員単位でのリサンプリングの記述については、前回答申の今後の課題として検討したことが残るような形にして、今後必ず実施するという趣旨にはならないようにすべき。

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>