

第18回 匿名データ部会 議事概要

1 日 時 平成26年12月19日（金） 13:00～14:45

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、加藤 久和、川口 大司

（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））、文部科学省、農林水産省、国土交通省、
東京都、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

国民生活基礎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しきい値基準に基づく上限値・下限値の設定

- ・畳数でも延べ床面積でもそうだが、豪邸のような家の特定化を防ぐためにトップコーディングのしきい値を下げたり、カテゴライズ化して提供したりするなど秘匿に配慮することも考えられるのではないかと。トップコーディングにおいてしきい値を設定する場合には、トップコーディングの対象となる属性の分布状況に基づいて、秘匿性と有用性の両方の観点から検討を行う必要があると考える。
- ・外観識別性からみた秘匿性の扱いは慎重に考えた方がよい。ただ、全国一本で扱うのであれば問題は無いのではないかと。
- ・住居の広さについて、大きな家と小さな家は識別されるのではないかとということだが、地域情報は存在しないので、特定できないのではないかと。
- ・トップコーディングとボトムコーディングを行う際に不詳が多数あった場合、どのように扱うのかが大変難しい。家計支出額のトップコーディングが平成19年調査と22年調査で大きく違うのは、どのように考えるべきなのか。
- ・家計支出額のトップコーディングは、現実のサンプルにより判断して行わざるを得ないので、平成19年調査と22年調査とで大きく違うのはやむを得ないと考える。過去のデー

タを比較する際には利用者側が工夫するしかないのではないか。標本の分布が母集団の分布とどれだけ合っているかについて、完全な母集団の情報はないかもしれないが、世帯人員の分布や世帯主の年齢別の分布などの基本的な情報は、母集団との対比が分かるようにしてほしい。

- ・サンプリングによって母集団が変わることもあり、社会経済状況の変化に伴い、トップコーディングの値は時勢を反映して動いていく。トップコーディングされたもとの平均値を示していただけると便利だと思うので、今後検討してほしい。
- ・匿名データと元の調査との分布に大差がないことを確認することが必要ではないか。

【部会長のまとめ】

- ・国民生活基礎調査は地域情報を出さない方向なので、1%ルールに従ってトップコーディングすることは妥当と判断する。

②平成22年調査において変更した匿名化措置

- ・手助けや見守りを要する者の状況で、主な介護者について、匿名性を担保するために「その他」を「その他の親族」に加えると、「その他の親族」の貴重な情報が埋没してしまうのではないかと。「その他」が特定できない以上、「その他」と「その他の親族」を一緒にする必要があるのか。
- ・「その他の親族」といった場合に、どのような間柄になっているか確認できているか。それは特定されやすい家族関係のものか。
→「その他の親族」の中身は確認していない。
- ・「その他の親族」も「その他」も、「その他」という言葉が入っていることで外観識別性は無くなっているのではないかと。
- ・調査票には「その他」と「その他の親族」は区別されているので、それをまとめて出すか、分けたままにするかだと思う。委員全体でまとめる必要がないという意見であれば、分けたままで良いのではないかと。
- ・こころの状態について、内訳を出すのは良い。今回の議論で、秘匿性を犠牲にせず有用性を高められるという議論が平成22年調査は基本的に認められたということだが、同じような議論が平成19年調査にも当てはまるのではないかと。厚生労働省の説明だと一度匿名データを作成して秘匿したものについては、将来も秘匿したままでいくということだが、このような前例を作るのは良くないのではないかと。
- ・そのような場合、遡及して直すのがベストだがリソースの問題もあり、実施する場合はリソースが許す範囲内での対応になるのではないかと。
- ・リソースの問題があるにしても遡及して直すのが良いと思うので、今回の例を一般化するものではないということを議事録に残した方が良い。

【部会長のまとめ】

- ・こころの状態については原データのままで提供するということとし、手助け見守りを要する者の状況における主な介護者のうち「その他の家族」と「その他」は分けて提供することを検討していただきたい。

③平成22年調査において新たに把握された項目の匿名化措置

- ・学歴については秘匿の観点から注意を払った方が良いのではないかと。「在学したことがない」人については、学歴と外観識別性の高い属性とのクロスによって特定化される可能性があることから、「在学したことがない」という区分に関しては、他の分類区分と統合した方が良いと考えられる。一方、高齢者の在学中の状況については、夜間中学に通っている方がいると思うが、その辺りの分布は気にしなくても良いか。全国一本の提供であれば、秘匿の観点から問題はないということになるのか。
- ・「在学したことがない」と回答した人と、「不詳」つまり回答しなかった人は、統計学的には区別した方が良い。
- ・「在学したことがない」人の外観識別性をどのように考えるか。相当高齢の方で、この状態などと比べると特定されてしまう可能性が高いと思われるので、全国一本であろうとなかろうと外観識別性を完全に否定しない方が良いのではないかと。
- ・外観識別の可能性はありうると考えた方が安心ではないか。「在学したことがない」の統合は、教育水準が一番低いところで行う方が良い。
- ・「在学したことがない」を「不詳」と統合すると、「在学したことがない」が無回答扱いとなるので、これは小学校中学校以下とみなして、最も低い学歴層に加えてグルーピングした方が良いのではないかと。
- ・匿名データは通常の集計結果とは違う扱いをしているということを明示すれば良いのではないかと。
- ・「在学したことがない」というのは、かなり強い意思表示ではないか。常識的に考えれば一番低い小学校中学校のところにカテゴライズした方が、不詳に入れるよりは良いと思う。

【部会長のまとめ】

- ・「在学したことがない」人の扱いを「不詳」に入れるのではなく、小学校中学校のところにカウント出来るかどうか、次回に回答いただきたい。

(2) 前回答申における「今後の課題」への対応

①何らかの地域情報を付与することの妥当性と可能性を検討することについて

- ・既に提供中のAデータとBデータに地域情報を付与するのは、これまで全国で扱われることを想定して匿名データを作成してきたことなどを考えると難しい。これとは別に、地域情報を提供することを主な目的として、できる限り支障のない変数に絞ったCデータを作成したとしても、情報量が少なく有用性を高めるかどうか不透明であるので、新しいCデータの作成について見送るとの判断については、やむを得ないのではないかと。
- ・地域情報を付与することによって失われる情報を考えると、付与しない方が良い。
- ・地域情報に対する一定のニーズはあると思うが、現行の匿名データの作成方法を保持する限りにおいては、地域情報の付与は難しい。

【部会長のまとめ】

- ・地域情報を付与することについては現段階では見送ることを妥当とする。

②所得等の内訳や世帯員別の情報の提供の妥当性と可能性について

- ・所得等の内訳を出せない理由として、合計値をトップコーディングしたものと、各要素をトップコーディングして合計したものが等しくならぬとのことだが、算術的には当然であり、理由になっていないのではないかと。失われる有用性を考えると、中身に年金や雇用保険や生活保護の話があり、どれも重要である。
- ・トップコーディングによる算術的な整合性が取れないとしても、所得等の内訳の提供することによる有用性は有るのではないかと。
- ・内訳の出し方は工夫の余地があるのではないかと。内訳と総計が一致しなくても構わないという考え方もあるし、秘匿の観点からセンシティブな情報が入っていることを念頭に置くのであれば、カテゴリー化して総計との整合性を気にしない形で提供するか、ノイズ等のパータベーションを用いるなど、いくつかの方法があるのではないかと。内訳の提供の可能性については、今後は是非検討してほしい。
- ・数値そのものは出せないとしても、どの所得が有るか無いかを出すだけでも分析上は違う効果が得られるのではないかと。金額のカテゴリーが出せないとしても、一番主な所得は何かとか、それ以外にこの所得も有るというフラグが立っているだけでもベターなので、提供の仕方を工夫してほしい。
- ・合計値をトップコーディングしたものと、各要素をトップコーディングして合計したものが等しくならぬことについて、利用者が混乱しないような説明が必要。秘匿性の問題を考えなければならないが、内訳をトップコーディングして提供することによる有用性が非常に大きいならば、費用対効果を考えても検討する価値があるのではないかと。ただ、世帯員票でクロスが出来てしまうと色々な属性が分かかってしまい、秘匿性の確保が難しくなるため、世帯員別に所得を出すことは個人的には反対である。
- ・内訳のある項目だけトップコーディングされても、他の数字が分かかってしまえば差し引きで出てしまうことが起こる。トップコーディングを一つひとつ行っていくのと同時に、一つ秘匿したときにもう一つ秘匿しなければならなくなり、それを判断するためのアルゴリズムは複雑になる。クロス統計表の秘匿と似たような問題だと思うが、二重の秘匿を行うことになるので、プログラムの作成者は相当苦勞することになるが、そこも含めて、実行可能性を検討してほしい。
- ・内訳を提供するためのプログラミングには相当な時間がかかることが予想され、今回提供できないのであれば、次回諮問時には内訳を出すことを検討すると答申案に記載してほしいかと。
- ・経済学的に考えれば、内訳が出た方が良くと思うので、どこまで出せるか検討してほしい。

(3) その他

- ・次回1月8日は、在学したことがない人の学卒の扱いについての再検討と、所得の内訳の匿名性を確保した提供方法について検討する。また、匿名データの提供時期の検討と全体のとりまとめをしたい。

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>