

匿名データ部会の審議状況について（報告）

第7回匿名データ部会 議事概要

1 日 時 平成23年6月7日(火) 10:00~12:11

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

樫広計部会長、井伊雅子部会長代理、宇賀克也委員、津谷典子委員、廣松毅委員、伊藤伸介専門委員、稲葉由之専門委員、安田聖専門委員、総務省(政策統括官(統計基準担当))、文部科学省、厚生労働省、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(総務省統計局統計調査部)】

高田調査企画課調査官、北林調査企画課課長補佐、横内調査企画課二次利用推進係長、高橋国勢統計課労働力人口統計室課長補佐

【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) 匿名データ部会の運営等について
(2) 労働力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

冒頭、部会長から新たに指名された稲葉専門委員の紹介が行われた後、次の議事が進められた。

(1) 匿名データ部会の運営等について

事務局から、参考1、参考2に基づき、匿名データ部会の公開方法及び開催日程について説明があった。

(2) 労働力調査に係る匿名データの作成について

まず、調査実施者から資料2に基づき、諮問37号「労働力調査に係る匿名データの作成について」の内容が説明された。続いて、事務局から資料3、資料4に基づき、本諮問に関する第45回統計委員会での委員からの意見及び黒田専門委員からの意見が紹介された。また、参考3に基づき、日本経済学会が2007年に公表した「政府統計匿名化データ利用に関する提言と要望」が紹介された。

更に、部会長から資料5に基づき、当該作成に関する論点案が示され、個別の論点に沿って、審議が行われた。各委員等の主な意見は次のとおり。

ア 地域区分に関する意見

- ・ 他の調査項目の匿名化措置は、地域区分を全国であることを前提に考えているようであるが、地域区分を全国からさらに細かくした場合、他の調査項目の匿名化措置の条件をどのよ

うに考えればよいか。

- ・ 以前、審議した全国消費実態調査では、本来6区分で提案されていた地域区分を、産業や職業などとの組合せが問題となったため、区分を粗くした経緯がある。
 - ・ 研究上の有用性を考えたときに、地域表彰をあまり粗くすると意味がなくなる。一方、産業分類が分かって、地域分類が分かってしまうと、匿名性の保持が非常に難しくなってくる。いろいろなものを掛け合わせたときにどうなるかを考えると、前回の4調査と同じように現段階では難しいのではないか。
 - ・ 本調査の集計でも、都道府県別の場合には四半期単位でしか公表していないにもかかわらず、匿名データで月次で地域とか都道府県別まで提供するというのは難しく、今回、全国一本というのは致し方がない。
- 地域区分について、概ねの意見は一致しているようであるが、次回以降、利用者の観点からの論点がないかについて、再度確認をしていきたい。

イ 同一年齢の子供の数に関する意見

- ・ 年齢階級の幅が0～3歳の4年刻みから13～14年までの2年刻みまで異なることや、0.5%基準の対象として高齢者のみの世帯も含んでいることなどを踏まえると、各年齢階層において一律0.5%基準を当てはめ、3人以上子供のいる世帯のレコードを全部削除してしまってもよいのか。
ワークライフバランスにおける保育サービスや母親の就業が問題となってくると思われるので、母親が20～49歳の再生産年齢の世帯を抜き出したときの子供の数と世帯数の関係を確認したい。
 - ・ ユーザーの目的にもよるが、例えば、0～6歳をまとめると削除対象を減らせる可能性があると思われるので、確認したい。
- 同一年齢の子供の数について、総務省には次回の部会までに「0～6歳」とした場合のレコード削除の考え方及び「20～49歳までの有配偶者女性の世帯」だけを抜き出した場合の子供の数と世帯数の関係を示した資料を作成していただきたい。その上で、概ねどのあたりのデータの切り方が妥当かを判断したい。

ウ 15歳以上の世帯員の年齢に関する意見

- ・ 85歳以上をトップコーディングすることについては、労働力調査であり、85歳以上の方が働いているということではないので妥当ではないか。
- ・ 5歳階級別のグルーピングに関して問題提起があったが、各歳刻みは大変厳しいと思う。2歳ごと、3歳ごとに分けることも考えられるが、極めて恣意的に思われる。どのように階級を分けるかについてももう少し議論すべき。

- ・ 人口学では、国際的にも5歳階級で切っていること、性年齢別の労働力率も5歳階級別に
出されているのでおそらく大丈夫ではないか。
- エ 月末1週間に仕事をした時間に関する意見
- ・ 1週間にどれだけ働いたかということは、外観識別性はそれほど高くないが、ごく少数の、
ある意味バイアスがかかってしまうような情報をトップコーディングすることは賛成であ
る。
- オ 事業の種類（産業）及び本人の仕事の種類（職業）に関する意見
- ・ 今回の産業分類は、特に製造業に関しては細かくなっているため、ここまで必要かとい
うのはあり得るが、匿名データの作成のために、これをいくつか統合して新たな分類を作成す
るのも、本調査の集計結果との関係で問題が起こりそうな気がするので、現時点ではやむを
得ないと考える。今回はこういう形で出してみても、ユーザーの方々からの意見を踏まえて、
今後少しずつ改良を加えていくというのが、現時点での考え方ではないか。
ただ、端的に大分類ではだめなのかとも考える。
 - ・ 実証分析を行う場合、回帰モデルの説明変数として産業を用いる場合は、大分類という形
でないとモデル分析は難しいのではないかと考える。しかし、産業構造を分析するために、
産業をより細かく見ることも考えられることから、産業分類については中分類でよいと思わ
れる。
- 産業分類等を見直した場合に、地域区分など他の項目で秘匿措置を緩和できるものはない
かという論点はあり得る。
- カ 前月欄の情報に関する意見
- ・ パネル化については検討されていないが、地域区分が全国一つということなので、できれ
ば産業や職業のデータを提供してもらえれば、労働力調査の重要な役割になると考える。こ
のため、前月欄の情報については、もう少し情報を加えていただきたい。現状では、集計デ
ータでできる以上の分析がかなり限られてくるのではないか。
 - ・ 本調査では、例えば、従業上の地位の異動を考えた場合に、従業上の地位と産業、あるい
は従業者数をクロスすることによって、特定化のリスクが高くなると思われる。このため、
前月欄の産業、従業上の地位、従業者数については提供しないとする今回の方針は妥当では
ないかと考える。
 - ・ ユーザーの多様な要望に応えるためには、例えば、匿名データとオーダーメイド集計とを
組み合わせた利用形態というのもあり得ると思う。少なくとも、今のところ前月欄の産業、
従業上の地位、従業者数を提供することはリスクが大きいと考える。
- キ その他の匿名化措置に関する意見

- ・ 死亡・転出のレコードを削除することについて、死亡・転出のあった世帯のレコードをすべて削除するのではなく、死亡・転出があったときの当該世帯員のレコードのみが削除されるという理解でよいか確認したい。

ク その他の意見

① 匿名化データの提供時期及び作成対象期間について

- ・ 特定化されるリスクはなるべく小さくするということを大原則としつつ、調査の性質によってある程度柔軟に運用するという観点からすると、今回、結果的に3年前のデータが提供できることは非常によいと思う。

② トップコーディングが行われた変数について

- ・ トップコーディング等が行われた変数の基本統計量については、速やかに提供していただくことが望ましい。

③ 同一世帯のマッチングについて

- ・ 例えばカナダでは、再度調査をし直して、きちんとマッチングしたパネルデータを作っている。これは、日本の匿名データとは異なるものであるため、棲み分けて議論を行ってほしい。確かにマッチングしたデータができればよいが、リスクが増えるので、匿名化するとほぼ使えないデータを作ることになる。海外でもデータは手に入るが、それは必ずしも匿名データという形ではないので、線引きをしていただきたい。

- ・ マッチングをして出すということは、匿名化という事業の中でやるよりは、完全に目的外の中でかなり厳しい規制の中で行うことが妥当ではないか。

- ・ イギリスでは、労働力調査のマイクロデータがパネルデータの形でも提供されている。イギリスの労働力調査の場合、パブリックユースのような形での提供はなされていないが、ライセンスを取得することによって、パネルデータの利用が可能になっている。こうした状況を踏まえると、わが国の匿名データにおいてもパネルデータの可能性に関する議論の余地はあってもよいと思う。

④ 集計用乗率について

- ・ 今回、リサンプリング率を8割としていることについては問題ないと思うが、乗率を使って実数値を出すとならば8割の数値になってしまう。実数値が推計人口に合っていた方が、分析の際に使い易いと考えられるので、集計用乗率の再付与について検討を行っていただきたい。

(3) その他

次回の匿名データ部会は7月4日（月）に開催することとなった。

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>