

匿名データ部会の審議状況について（報告）

第10回 匿名データ部会 議事概要

1 日 時 平成24年11月15日(木) 10:00~12:06

2 場 所 中央合同庁舎4号館4階 共用第4特別会議室

3 出席者

(部会長) 椿広計

(委員) 安部由起子、津谷典子

(専門委員) 加藤久和、安田聖

(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、厚生労働省、農林水産省、
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: 井上調査企画課長、岩佐国勢統計課長ほか

(事務局) 内閣府統計委員会担当室: 村上室長、若林参事官ほか

4 議 事

(1) 匿名データ部会の運営等について

(2) 国勢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

部会長から部会長代理として津谷委員が指名された後、以下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3、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) 国勢調査に係る匿名データの作成について

まず、総務省統計局から資料1に基づき、諮問44号「国勢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。続いて事務局から参考4に基づき、本諮問に関する第58回統計委員会での委員からの意見が紹介された。更に、資料2に基づき、部会長による当該匿名データの作成に関する論点(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 地域区分について

- ・ 後の年齢の議論とも関わってくるが、世帯員の年齢を5歳階級区分とすることについては、例えば東京都では各歳別で提供するなど、都道府県や大きな市などで違う取扱をすることはできないか。

→ 全国一律の基準を設けてデータを提供する方がよいと考えている。ただし、世帯人員のように地域別の差が大きいものについては、一律の基準だと削除対象が多くなるため、例外的な措置として地域別の基準を採用している。

- ・ 地域区分を「都道府県」及び「人口50万以上の市区」とすることについては、有用性の観点

から妥当であることを確認する一方で、各歳別の情報を提供するためには、例えば地域区分なしの匿名データも考えられるので、その必要性を今後議論していくことが考えられる。

イ サンプリングの方法等について

- ・ 「一般世帯」は世帯単位、「施設等の世帯」は個人単位と、サンプリングの単位が異なっているにもかかわらず、同じ1%を抽出しているため、代表性に歪みが無いかを確認しておく必要がある。
- ・ 抽出単位の異なるサンプルを統合しているので、世帯の種類ごとの抽出率など、ウエイトに近い情報を提供するか、あるいは、ユーザーが公表値から自分で作成するように注意を喚起するかのどちらかではないかと思う。サンプリングとしては、これ以上の方法は無いと思う。
- ・ 諸外国の人口センサスのパブリック・ユース・サンプルでは、ウエイトを提供している場合が多く、世帯と個人のウエイトを別々に提供している場合もある。ウエイトを匿名データに付けることが有用ということであれば、それをできるかどうかということだと思う。将来的には、日本だけでなく世界中のユーザーに利用していただいた方がよいので、諸外国で既に提供されているパブリック・ユース・サンプルの方法にある程度従った方がよいのではないか。
- ・ ウエイトがバイアスの要因となる場合もあり、使い慣れていない人にとっては問題になる可能性がある。ウエイトをユーザーに提供する場合にどのように説明していくかということもあり、コストとベネフィットがあると思う。
- ・ 集計値としての利用であれば、あまり問題にならないかもしれないが、推計して分散を評価するとすると、サンプリングの部分が効いてくる。ウエイトの公表の可能性については、公表に関わる問題ということで、追加の論点として議論したい。

ウ 情報の削除について

(ア) 年齢差の大きい夫婦のいる世帯

- ・ 年齢差が25歳以上の夫婦のいる世帯を削除することについては、なぜ25歳以上なのか。
→ 地域ごとの分布を見て、年齢を5歳階級区分で提供することもあり、25歳以上と設定している。

(イ) 父子世帯

- ・ 父子世帯の都道府県ごとの割合は公表資料でわかるのか。
→ 国勢調査では父子世帯及び母子世帯に関する統計表を公表しており、既存の公表資料からわかる。
- ・ 今回、地域の情報を提供する一方でスワッピングも導入しており、削除の基準を緩和できる場合がある。その辺りを加味して議論する必要がある。

(ウ) 年齢差の大きい又は小さい親子のいる世帯

- ・ 年齢差の大きい又は小さい親子のいる世帯は、再婚等で多くなっている。年齢差を地域別に実数で見た場合に、削除の基準を緩和できないか検討の余地がある。
- ・ 計画されている削除の基準が、どの程度まで許容したことになっているかについて確認したい。

エ 世帯員及び世帯に関する項目の再編等について

(ア) 年齢

- ・ 議論の前提として、調査票情報の特性は統計調査ごとに異なることから、ガイドラインにおいても、各統計調査について一律に匿名化の基準を設定するのではなく、個々に検討し設定する必

要があるものとされている。また、調査年次によっても分布が変わるので、トップコーディングを行う年齢区分についてはその都度考えていただきたい。

- ・ 日本全体で高齢化が進んでいる中で、地域別に見ると非常に高齢化が進んでいる県とそうでない県の差が非常に大きくなってきている。今回、85歳以上をトップコーディングすることが適切か、検討する必要がある。
- ・ 寿命を考えると、85歳以上でトップコーディングするのは若過ぎる気がするので、見直す必要があるのではないか。
- ・ 85歳以上については、他の属性情報と合わせた場合に特定されないか、詳細に議論する必要がある。

(イ) 産業（大分類）及び職業（大分類）

- ・ 産業について、「製造業」と「電気・ガス・熱供給・水道業」の統合は、どのようなメリットとデメリットがあるかを考えておく必要がある。また、職業について、「保安職業従事者」、「農林漁業作業員」及び「運輸・通信従事者」を統合することについては違和感がある。特に通信が増えていることを考えると、ここは議論しておいた方がよい。
- ・ 「鉱業」と「電気・ガス・熱供給・水道業」は「その他」に統合した方が有用だと思う。ユーザーとしては「建設業」や「製造業」を見たいので、そこに他のものが混ざっていると違和感がある。

(ウ) 5年前の住居の所在地

- ・ 平成12年の国勢調査の設問だが、具体的に5年前に住んでいた都道府県名は出さないということか。
- 「他の市区町村」と「外国」とを統合する計画であるため、「他の市区町村」の内訳である5年前に住んでいた都道府県名等は提供できないものとなる。

(エ) 従業上の地位

- ・ 「雇人のある業主」、「雇人のない業主」及び「家庭内職者」を統合することについて、自営業主に家庭内職者を入れるのは悩ましいところである。
- 「家庭内職者」をどこに統合するかであるが、定義からは「家族従業者」は雇われている者に近く、むしろ「家庭内職者」は自営業主に近いと考えられるため、このように統合する計画である。なお、報告書においても、このような考えから、自営業主と「家庭内職者」を統合し、「自営業主」として表章しているところでもある。
- ・ 先ほどから、項目の統合に関していくつか意見が出されているが、ユーザーの立場で合理的であると思われる統合の仕方をした場合に、どのようなデータとなるのかを確認しておくということではないかと思う。また、それが匿名化の観点で問題ないかということも議論する必要がある。分析者の観点からこのような統合の方がよいのではないかという方法があれば、それを基に次回以降の部会で検討することでよい。

(オ) 家計の収入の種類

- ・ 「家計の収入の種類」については、高齢者の収入が年金かどうかということは重要な情報であるので、統合等をするにはあるにしても、情報としてはあった方がよい。
- 「家計の収入の種類」については、指摘を受けて、どのような方法がよいかを検討しているところであり、次回には何らかの草案を提示できるのではないか。
- ・ 「家計の収入の種類」については、仮に匿名化して提供することが必要という判断をした場合には、どのような匿名化が適切であるかを含めて、次回以降検討したい。

オ スワッピングについて

- ・ スワッピングについては、諸外国の人口センサスのパブリック・ユース・サンプルでは通常用いられている。地域の情報を提供する以上、スワッピングは必要である。
- ・ スワッピングを行うことに異議はないが、実際に基本統計量のようなもので確認しておいた方がよいのではないか。ユーザーに確認できる情報を提供していただきたい。
- ・ 集計量に歪みが出るようなスワッピングでは有用性の観点から問題になる。逆に言えば、サンプリングの誤差と比較して無視できる程度の影響であることが確認できればよいのではないか。この部会や委員会で、分布に違いが無いことを保証していくということかと思う。
- ・ スワッピングは賛成である。ただし、集計値を見たいのであればあまり影響はないと思うが、より詳細な分析をしたい場合もあるので、その辺りの情報提供をしていく方が有益ではないか。
- ・ スワッピングに関する情報が少ないと、専門誌のレフェリーやエディターが危惧するかもしれない。
- ・ アメリカのパブリック・ユース・サンプルと日本の匿名データとは必ずしも一致していないので、一概に同じ議論をするのは難しいかもしれないが、アメリカのパブリック・ユース・サンプルで書いた論文は受理されないと聞いている。州単位での分析は認められるが、それ以上については個票で分析すべきであるということを宣言している。
- ・ 国勢調査であればオーダーメイド集計も利用できるので、匿名データとオーダーメイド集計をうまく組み合わせて、学会からも認められるようにするという事ではないか。
- ・ スワッピングに関しては、一般のユーザーがどのように認識するかという問題もあるので、この部会でどのような説明の仕方をするのが妥当かを議論しておきたい。

カ その他の意見

- ・ トップコーディングの基本統計量は、将来提供されることを期待している。また、匿名データの提供時期については、実査への影響を避けるために、一つ前の調査までを提供することを前提に考えていたと思う。今回の国勢調査の周期は5年であるが、他の調査では変わってくるので、5年に縛られずに検討すればよいのではないか。

(3) その他

次回の匿名データ部会は、12月7日（金）10時から中央合同庁舎第4号館 共用第3特別会議室で開催することとされた。

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>