

第2回匿名データ部会 議事概要

1 日 時 平成21年2月13日(金) 10:00~12:00

2 場 所 中央合同庁舎第4号館2階 共用第3特別会議室

3 出席者

廣松部会長、井伊部会長代理、宇賀臨時委員、椿臨時委員、津谷臨時委員、玄田専門委員、西郷専門委員、永瀬専門委員、星野専門委員、安田専門委員

総務省(政策統括官室(統計基準担当))、厚生労働省、経済産業省、国土交通省、東京都、千葉県、日本銀行

【諮問者】

杉山総務省統計局統計調査部調査企画課長、栗原総務省統計局統計調査部調査企画課調査官

【事務局】

中島内閣府大臣官房統計委員会担当室長、高木内閣府大臣官房統計委員会担当室参事官

4 議事次第 (1) 全国消費実態調査、社会生活基本調査、就業構造基本調査及び住宅・土地統計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

(1) 全国消費実態調査、社会生活基本調査、就業構造基本調査及び住宅・土地統計調査に係る匿名データの作成について

「配付資料」の部会長論点メモにおける個別の論点について、以下のとおり、諮問者である総務省統計局から説明を受けた後に審議が行われた。

ア リサンプリングの方法

《総務省統計局の説明》

リサンプリング率については、作成計画案の抽出率により仮作成した匿名データによる統計と全標本から作成している公表統計との間で、代表的な項目に関し平均値等を比較した結果、乖離が少ないことが確認できたこと等から、適当であると考えます。

世帯単位によるサブサンプルの抽出については、過去の研究においても世帯関係に着目した分析が多く、利用者ニーズが高いと考えられることから、個人単位によるサブサンプルの抽出よりも適当なものであると考えます。

《各委員の主な意見》

- ・ 世帯人員8人以上の世帯のレコードについては、匿名性を確保するため、匿名データから削除することとしているが、個人単位によるサブサンプルの抽出を導入することにより、一部でもデータを残すことを検討すべきではないか。

- ・ 過去の一橋大学における研究例から見ると、世帯人員 8 人以上の世帯について、個人単位によるサブサンプルの抽出を導入しても、個人単位のみ情報は利用されない可能性が高く、当該導入の必要性は乏しい。
- ・ 標本の各レコードが持つ集計用乗率と抽出地域に一定の対応関係がある場合、当該乗率から抽出地域が特定されることを防ぐため、レコードを乗率別に層化した上で各層から乗率の大きさに応じて確率比例抽出を行い、その層の平均乗数を再付与する措置を講じているが、層化していれば層内の乗率にあまり違いがないことから、確率比例抽出まで行う必要はないのではないか。
- ・ リサンプリング率及びサブサンプルの抽出単位については、匿名性及び有用性の確保の観点から見て、作成計画案の考え方で問題はない。

イ 地域区分

《総務省統計局の説明》

地域区分を全国 6 ブロック別とした「全国消費実態調査」等 3 調査については、前回部会の指摘を受けて検討した結果、公表統計との照合により都道府県の別が明らかになる事例が一部あることが判明したため、「3 大都市圏」及び「その他の地域」の 2 区分にすることとしたい。

一方、住宅・土地統計調査の場合、世帯・個人等に関する情報が相対的に少ないことを勘案し、都道府県別とする。

《各委員の主な意見》

- ・ 3 調査について 2 区分とすることは保守的であると思うものの、匿名性の確保の重要性を勘案すると、今回はやむを得ない。
- ・ 3 調査における地域区分の細分化は、将来の検討課題である。

ウ 個人の年齢

《総務省統計局の説明》

年齢については、調査客体の特定リスクが大きいため、諸外国の例も勘案し、15 歳以上の者は 5 歳階級別としたい。

年齢に係るトップコーディングの上限値については、前回部会での指摘を受け、4 調査とも一律 85 歳以上としたい。

15 歳未満の若年者については、調査客体の特定に利用できる属性情報が限定されていることから、各歳別に提供することとする。

《各委員の主な意見》

- ・ 就業構造基本調査については、年齢を各歳別にしても、調査客体が特定されるリスクは少ないのではないかと。労働分野の研究においては、詳細な職業分類に関する研究よりも年齢を中心とした研究の方が主流なので、職業分類を 10 分類程度とする一方で、年齢を各歳別に提供して欲しい。
- ・ 労働分野の研究実績から見て、就業構造基本調査の匿名データにおいて各歳別に提供する必要性は低いのではないかと。
- ・ 年齢を各歳別にする一方、職業等の分類区分を大幅に粗くしたデータセットをもう一つ作ることはできないかと。

- ・ データセットを2つ作ると、これらの結合によってリサンプリングの匿名化効果が失われる危険性が生じ、安全性が低下する。
- ・ 15歳以上の者の年齢区分を5歳階級別とすることについては、匿名データの作成が今回初めてであることや調査客体の匿名性の確保の重要性等を勘案すると、やむを得ない。
- ・ 現時点では、年齢に係るトップコーディングの上限値を4調査とも一律85歳以上とすることは適当だが、近年の高齢化の進展状況等を勘案すると、将来、当該上限値の見直しを行う必要がある。

エ 世帯人員、床面積、家賃等

《総務省統計局の説明》

世帯人員が8人以上の世帯については、当該世帯レコードの中の他の情報との関係からトップコーディングが困難であり、当該世帯レコードそのものを削除することが必要である。

床面積等については、地域間の比較を考えた場合、都道府県毎にばらばらの上限値を設定しても有効な比較が出来ないこと等から、一律にトップ（ボトム）コーディングの上（下）限値を設定する。

トップコーディングを行った場合の平均値等の提供については、運用後のニーズ等の状況を踏まえ、検討することとしたい。

《各委員の主な意見》

- ・ 世帯人員が8人以上の世帯を削除することは適当だが、近年の世帯規模の縮小傾向等を勘案すると、将来、削除すべき世帯の世帯員数について見直しを行う必要がある。
- ・ 家賃等のトップ（ボトム）コーディングの上限値が全国一律でないとは有効な地域比較ができないとする説明には疑問がある。
- ・ 家賃等のトップコーディングの上限値を9万円とすることは、都市経済学分野で大都市圏の家賃分析を行っている研究者にとっては、匿名データの有用性の低下につながるのではないか。
- ・ 床面積、家賃等におけるトップ（ボトム）コーディングの上（下）限値については、匿名データの作成が今回初めてであること等を勘案すると、やむを得ない。

オ 全国消費実態調査における年収等が高額な世帯

《総務省統計局の説明》

当初、年収等が高額な世帯は匿名データから削除することとしていたが、前回部会での議論を踏まえ、その取扱いを再検討した結果、当初案どおり削除（案1） 0.5%基準に従って削除（案2の1） 提供する項目を限定し0.5%基準によるトップコーディングを実施（案2の2） 資産関係項目を活かし関連する他の情報を削除（案3） の4つの見直し案のいずれかにすることとしたい。

《各委員の主な意見》

- ・ 匿名データの信頼性の確保の観点から、サブサンプルの削除は必要最小限に留めるべきであること等を総合的に勘案し、案2の2が適当である。
- ・ 全国消費実態調査は、年収等に関する具体的な数字が出ており、また、他の統計調査以上に

調査客体との信頼関係で成り立っている調査なので、匿名性の確保を重視すべき。

カ 匿名データの作成対象年次

《総務省統計局の説明》

提供する匿名データについては、調査客体の匿名性の確保の観点から、平成以降の調査で、かつ調査実施から5年以上経過したものとしているが、今後、平成より前に実施したものも提供できるように検討したい。

《各委員の主な意見》

- ・ 平成より前に実施した調査のみならず、直近の調査についても今後検討すべき課題である。

キ その他（匿名化措置の考え方等）

《総務省統計局の説明》

0.5%基準は、諸外国においても、トップコーディングに関する基準として適用しており、全ての情報について適用している訳ではない。

前回部会において学生等のためにハードルを下げたデータを作るよりも訓練する機会等を提供することが重要といった意見があったことや、レコード同士の照合により安全面での問題があるとの意見もあったことから、複数の匿名データを作成することは適切ではないと考える。

《各委員の主な意見》

- ・ 全ての変数に0.5%基準を適用する必要はないが、極めて珍しい職業や外から見て分かる変数については、調査客体の特定リスクの有無を確認した方が良い。
- ・ 最終的に匿名データを提供する段階で、該当する個体数が極端に少ないものの有無を確認し、その結果、調査客体が特定される危険がある場合には、必要な処置を行うべきである。

(2) その他

今回の匿名データ部会は2月24日（火）に開催することとなった。

<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>