

匿名データ作成のこれまでの取組 及び今後の検討について（案）

令和4年9月16日 統計研究研修所

これまでの取組

公的統計基本計画（平成30年3月6日閣議決定）

匿名データについて、**統計研究研修所の支援を受け**、より広い範囲の者が利用できるようにする形での提供に向け、必要な法制面、**技術面からの検討**を踏まえ、**早期の提供を検討**する。

統計研究研修所における対応

令和3年度までに提供可能な直近の年次の検討が終了

- 統計局所管の6調査及び厚生労働省所管の1調査
（厚生労働省の賃金構造基本統計調査については、新規作成の検討において技術的な支援を行った）
- 「匿名化処理基準」に基づき、匿名データ有識者会議において検討

作成方法を一元的に確認 ⇒ 課題が散見

[匿名データ有識者会議での検討実績]

調査名	作成方法 WG	有識者会議
国勢調査（H22,27）	R1年8月	R1年10月
社会生活基本調査（H23,28）	R2年2月	R2年7月
就業構造基本調査（H24,29）		
全国消費実態調査（H21,26）	R2年 8月,10月	R2年11月
労働力調査（H25～R1）		
国民生活基礎調査（H28）		
住宅・土地統計調査（H30）	R3年 6月,8月	R3年9月

匿名データの作成及び利用における課題

1. 現行の匿名化処理の方法

(1) 「匿名化処理基準」の記載内容

- ✓ 同じ調査項目の処理方法の調査間の記載内容が異なる
⇒ 考え方の統一性がなく、検討作業の負担が大きい
(加えて、処理方法の秘匿のため公開する資料の内容についても課題)

(2) 匿名化処理された項目の有用性

- ✓ 安全性を重視して分類区分数を少なくしている
- ✓ 調査客体の属性等が考慮されておらず、一律のしきい値を適用している
(連続変数(金額、面積など))
⇒ 分析の有用性に課題

2. 利用環境に課せられる管理要件

- ✓ 利用者が用意する環境は、調査票情報の利用の際とほぼ同じ管理要件が必要
⇒ ハードルが高い

【課題の検討】 1. 現行の匿名化処理の方法

【課題(1)】 「匿名化処理基準」の各調査での記載内容に差異

- 調査によらない一般的なルールの整理

- 匿名化を考慮すべき調査項目の整理
- 調査項目の内容や性質（連続変数、離散変数等）により処理を類型化
- 特異な世帯（世帯人員が多い、同一年齢の子供が多い等）の処理方法の見直し
- 調査（集計）の分類区分別の度数の大小のみによらない匿名化処理のルール化
- 同一処理方法について表記を統一

⇒ 一般的なルールに準拠することにより、検討作業の効率化を図る

【課題(2)】 一部の調査項目において分析の有用性に課題

- 現行の匿名化レベルを維持しつつ、有用性に配慮した処理方法をルール化

- 匿名化された調査項目の分類区分の細分化の見直しを検討
- 調査において特徴のある調査項目の分類区分を考慮したしきい値の設定の検討
- 同一調査、同一年次における複数ファイルの作成方法の検討
- 上記の検討において、既に提供している匿名データとの分析の継続性に留意

⇒ 有用性に配慮したルールに基づき、利用しやすいデータの作成を目指す

 **調査横断的かつ一般的なルールの整理**

【課題の検討】 2. 利用環境に課せられる管理要件

【課題】 利用時に厳しいデータの管理要件が課せられている
(利用場所限定、管理方法の徹底など)

- 攪乱処理のさらなる拡充（現状は、一部調査に適用）
 - 匿名性を高め、さらなる安全性を確保
 - 管理要件を緩和し、利用しやすい環境へ
 - 処理後において、安全性が確保されているか定量的な確認方法の検討

⇒ 安全性と有用性を両立させ、利用環境の改善を図る

データの作成の技術的な検討

今後について

これらの課題については、共通課題検討WGで検討し、その内容を踏まえ、国勢調査（令和2年）で作成方法WGで検討及び試作

※ 特に攪乱処理の拡充については、匿名化処理に関する共同研究等の成果を踏まえ検討を進める

匿名データ作成に係る検討スケジュール

★：作成方法WG ☆：共通課題WG ◎：有識者会議 ◆：提供 ※：統計委員会

