

別紙様式第 2 号

匿名データの作成方針【雛形】（案）

－初めて匿名データの作成を行う統計調査の場合－

1 基本的な考え方

本調査の匿名データ化については、「匿名データの作成・提供に係るガイドライン」（平成〇〇年〇月〇日付、総務省政策統括官（統計基準担当）決定）の匿名化処理基準を参考としつつ、本調査の特徴を踏まえて所要の秘匿措置を講じる。

2 作成する匿名データの構成概要

本調査に対する利用ニーズ、調査体系の特性を活かし、世帯又は個人単位の以下の匿名データを作成する。

調査本体の 標本の大きさ	リサンプリング率	匿名データの 標本の大きさ
約〇〇万世帯	〇〇%	約〇万世帯

3 適用する匿名化処理

本調査では、調査票情報に対して、以下の匿名化処理を適用する。

(1) リサンプリング

匿名化処理基準を踏まえ、リサンプリング率〇〇%を目安としてリサンプリングを行う。リサンプリングは、《リサンプリングの方法を記載（例：地域区分による層化を行ったのち、世帯を単位としてまとめた上で等確率抽出により、世帯を単位とするリサンプリング率が約〇〇%になるようにする。）》

(2) 識別情報

以下の秘匿措置を講じるほか、出現数が少なく個体識別リスクが考えられるレコードは削除する。（詳細は、別添「匿名データの作成に係る審査表」を参照）。

例)

ア 世帯人員

世帯人員が多数である世帯のレコードは削除する。

イ 地域区分

例 1) 地域区分は「全国」のみとする。

例 2) 地域区分は匿名化処理基準を踏まえ「3大都市」と「3大都市以外」の2区分とする。

例 3) 地域区分は都道府県とする。 等

ウ 年齢

個人の年齢は、15歳未満は各歳、15歳以上は5歳階級でグルーピングするとともに、高齢者についてはしきい値基準に基づきトップコーディングを行う。

エ 住居の床面積

所得総額等は世帯総額のみをトップコーディングし、その内訳情報は削除する。

オ その他

これら以外にも、リスクを低減するために、レコードの削除、トップコーディング、ボトムコーディング、リコーディング、乱数によるレコード順の並び替え等、必要な措置を行う。

また、トップコーディング、ボトムコーディング、リコーディングに当たっては、利便性を考慮すると共に、統計調査の本体集計の結果表章に用いられる分類を参考とする。

匿名データの作成方針【雛形】

－匿名データの年次追加を行う統計調査の場合－

1 基本的な考え方

本調査の匿名データ化については、「匿名データの作成・提供に係るガイドライン」（平成〇〇年〇月〇日付総務省政策統括官（統計基準担当）決定。以下、「ガイドライン」という。）の匿名化処理基準に準拠した秘匿措置を講じて、作成、提供する。ただし、社会情勢の変化や他調査の作成方法等を勘案し、当該年次の特性に応じた措置を講じる。

2 作成する匿名データの構成概要

今まで作成を行ってきた〇〇〇〇調査に係る匿名データと同様に、以下の匿名データを作成する。

調査本体の 標本の大きさ	リサンプリング率	匿名データの 標本の大きさ
約〇〇〇万世帯	〇〇%	約〇万世帯

3 適用する匿名化処理

本調査では、ガイドラインの匿名化処理基準に準拠した匿名化処理を適用する。
なお、新規の調査項目及び社会情勢の変化等による変更点は以下のとおりである。

(1) 新規の調査項目

- ① 就業期間：50年以上は50年でトップコーディング（1%のしきい値基準）

(2) 社会情勢の変化等

- ① 世帯人員が多い世帯の削除

8人以上の世帯の割合が減少したため、9人以上を8人以上に引き下げ（しきい値基準）

- ② 年齢のトップコーディング

人口高齢化により85～89歳の割合が増加したため、85歳以上を90歳以上に引き上げ（しきい値基準）

- ③ 年収額のトップコーディング

1200万円以上の割合が減少したため、1200万円以上を1100万円以上に引き下げ（しきい値基準）