参考2

匿名データの作成に係る匿名化処理基準

Ⅰ　統計調査共通で適用する処理（※しきい値は、各調査において設定）

1　識別情報に関する処理

(1) 提供しない調査事項等

調査対象を特定する危険性の高い識別情報である調査事項等（氏名、住所、出生の年月、調査地域の番号等）は提供しない。

ア　個人が特定できる調査事項等

- 氏名、勤め先 など
- 出生の元号・年・月（提供する場合は、年齢に換算する）

イ　調査対象が特定できる調査事項等

- 調査地域に関する情報（実査で使用した調査区番号などの市区町村よりも詳細な地域情報）、建物が特定される調査事項 など

ウ　現住居以外の地域に関する調査事項等

- 以前住んでいた場所、従業している場所、通学している場所 など（市区町村などの地域情報）

エ　直近の災害等に関する調査事項

(2) 調査対象の削除

調査票情報において、提供する地域ごとの出現頻度が低く特定される可能性の高い調査対象を削除することを検討する。

ア　世帯人員に関する削除処理

- 世帯人員について、累積の構成割合又は度数が、設定したしきい値により決定した上限値を超えた人数のいる世帯（直近の国勢調査のしきい値を確認し検討する。）

イ　同一年齢の世帯人員に関する削除処理

- 同一年齢の世帯人員について、累積の構成割合又は度数が、設定したしきい値により決定した上限値を超えた人数のいる世帯（原則として、15歳未満を対象とする。また、直近の国勢調査のしきい値を確認し検討する。）

ウ　介護等に関する調査事項等に該当する世帯人員についての削除処理

- 介護等に関する調査事項等に該当する世帯人員について、累積の構成割合又は度数が、設定したしきい値により決定した上限値を超えた人数のいる世帯

(3) 実査の際に付与される情報の匿名化処理

実査の際に付与される情報について、調査対象が特定される可能性が高い場合、匿名化処理を行うことを検討する。

ア　調査票情報の配列順の並べ替え（調査対象のランダムソート）

- 調査対象が特定できる可能性の高い識別情報（世帯番号、住宅番号等が該当）は、配列順の並べ替え、再付与等を検討する。

イ　集計用乗率（復元倍率）の再付与

- 集計用乗率（復元倍率）から調査対象の抽出方法が判明し、調査対象が特定される可能性が高くなる場合は、有用性を考慮した上で再付与等を検討する。

2　調査事項等の内容や性質によりしきい値を設定し行う処理

提供する地域等ごとの調査事項等の分布状況において、設定したしきい値よりも出現する構成割合又は度数が少ないなど調査対象が特定される可能性が高くなる調査事項等について、内容や性質と統計調査の目的との関係性を考慮して、匿名化処理を検討する。
なお、調査事項に回答する調査対象が限定されるなど、調査対象が特定される可能性が高くなる場合は、調査事項等を提供しない 又は 調査対象を削除することを検討する。

(1) 量的データに関する調査事項等（数量を直接記入又は記入内容から数量に換算する調査事項等）

（年齢、階数、面積、回数、時期（〇年〇月）、期間（〇年〇か月）、時間、金額などが該当）

ア　累積の構成割合又は度数の分布状況から、設定したしきい値により決定した上限（下限）値を上（下）回る場合には、トップ（ボトム）コーディング

- 上限（下限）値は、提供する地域、世帯の種類、建て方などの区分別に検討する。
- 同一の調査事項等について、調査票ごとに上限（下限）値が異なる場合は、より粗い値を採用する。
- 単位にも留意する。

イ　構成割合又は度数の分布状況により、リコーディング（階級区分）

- 本体集計等で使用している区分、有用性を考慮し、階級区分を検討する。

ウ　他の調査事項等の匿名化処理に伴い匿名化処理が必要となる場合、上記ア、イ等の匿名化処理

(2) 質的データに関する調査事項等（「(1)量的データに関する調査事項等」以外）

ア　構成割合又は度数の分布状況により、設定したしきい値により決定した下限値を下回る場合には、リコーディング

- しきい値は、提供する地域、世帯の種類、建て方などの区分別に検討する。
- 「不詳」又は 不詳に相当する区分は、リコーディングしない。
- 「その他」は、原則としてリコーディングしない。
  （ただし、「その他」以外の区分にリコーディングの対象となる区分がない場合は、「その他」とリコーディングすることは可）
- 区分を統合する場合は、その内容及び性質、本体集計等で使用している区分、有用性を考慮しリコーディングを検討する。
- 同一の調査事項等について、調査票ごとに区分等が異なる場合は、より粗い区分を採用する。
- 介護、教育等に関する内容の調査事項等について、他の調査事項等との関連性を確認した上でリコーディングを検討する。
- 調査事項等が、意識を問う内容（「なぜですか」、「～したいですか」などが該当）の場合、そのまま提供できるかを検討する。

イ　他の調査事項等の匿名化処理に伴い匿名化処理が必要となる場合、上記ア等の匿名化処理

Ⅱ　各統計調査で独自に行う処理

1　提供する地域（市区町村以上の地域情報）

各統計調査において、提供する地域（都道府県、市区町村等）を決定する。

2　サンプリング・リサンプリング率

各統計調査において、調査方法を考慮し抽出した調査票情報の一部を提供する。

【国勢調査】	一般世帯は世帯単位に、施設等世帯は個人単位に約１％
【社会生活基本調査】	世帯単位に約80% （調査票Ａ、調査票Ｂごとに抽出する。）
【就業構造基本調査】	世帯単位に約80%
【住宅・土地統計調査】	住宅単位に約10% （調査票甲及び乙を合わせて抽出し、両方の調査票に共通の調査事項等を提供する。）
【労働力調査】	世帯単位に約80%（基礎調査票を抽出する。但し、沖縄県は約20%とする。）
【全国家計構造調査】	世帯単位に約80% （次の世帯について、世帯の種類（二人以上の世帯、単身世帯）ごとに抽出する。 ・基本調査世帯、家計調査世帯特別調査世帯及び全国単身世帯収支実態調査世帯 ・簡易調査世帯 世帯票、家計簿、年収・貯蓄等調査票がすべて揃っている世帯を提供する。）
【国民生活基礎調査】	世帯単位に約20%
【賃金構造基本統計調査】	労働者単位に約40%

3　世帯・個人識別情報の匿名化

各統計調査の特性により調査対象が特定される可能性が高い場合は匿名化処理を行う。（提供する地域ごとに検討を行う。）

【国勢調査】	・母集団に対して一意又は二意となる世帯又は個人がいる世帯の削除 ・父子世帯の削除 ・子供の数が多く、世帯主・配偶者が外国人である世帯を削除 ・年齢差の大きい夫婦のいる世帯を削除 ・年齢差の大きい又は小さい親子のいる世帯を削除 ・15歳未満の就業者のいる世帯を削除
【社会生活基本調査】	・母子世帯、父子世帯において、出現頻度の少ない世帯を削除
【住宅・土地統計調査】	・家計を支える者の年齢が15歳未満の世帯の削除
【労働力調査】	・自衛官及び受刑者のレコードを削除
【国民生活基礎調査】	・父子世帯の削除 ・年齢差の大きい夫婦のいる世帯を削除 ・年齢差の大きい又は小さい親子のいる世帯を削除
【賃金構造基本統計調査】	・公営の事業所の削除

4　撹乱処理

調査単位、調査対象等が、特定又は推定されないようスワッピング、誤差の導入などの匿名化処理を必要に応じて行う。

【国勢調査】	・２つの調査対象を入れ替える。（スワッピング）