

## 参考 匿名化技法について

一般国民からの信頼と協力により集められた調査票情報を基に作成する匿名データは、学術研究及び教育目的のため広く一般に提供されることから、被調査者が特定できないよう加工することが統計法で規定されています。

具体的には、単に氏名・住所を削除しただけでは不十分であるため、匿名データの作成に際しては、以下の匿名化技法により、被調査者の情報を確実に秘匿します。

### 1 情報の削除

#### (1) データの再抽出（リサンプリング）

元の統計調査のデータすべてを匿名データに用いるのではなく、そのうちの何割かを再抽出したデータを用いる。

#### (2) 直接的な識別情報の削除等

直接的な識別情報は、データから削除する。（例：地域を削除）

また、データの配列順により特定されないように、無作為に並べ替えを行う。

#### (3) 裾切りによるデータ削除

特徴的で、出現率が低い値があるデータは、削除する

（例：多人数世帯、3つ子以上世帯を削除）

### 2 識別情報の階級区分統合

#### (1) 上限（下限）階級区分の統合（トップコーディング、ボトムコーディング）

極端に大きな（小さな）値は、上限（下限）値を設けて統合する。

（例：一定の年齢以上を統合、所得の一定金額以上を統合）

#### (2) 再コード化（リコーディング）

分類事項の程度は、詳細なものではなく、粗いものとする

（例：各歳階級を5歳階級化、細かい分類を粗く再分類化）