

「社会生活基本調査に係る匿名データ作成について」の論点

平成26年10月3日

1. 匿名性及び有用性の確保

社会生活基本調査（調査票B）（平成13年及び18年）の匿名データ作成方法について、社会生活基本調査（調査票B）が他の調査に比べ小規模標本であることや詳細な生活行動分類符号が付与されていることなどの特性を踏まえ、以下の事項の観点から匿名性及び有用性が確保されているか。

(1) 匿名化措置をしている事項

ア ファイルの種類

世帯単位のファイルとしているが適切か。

イ 地域区分

地域区分を「全国」としているが適切か。

ウ リサンプリングの方法

リサンプリングは、世帯を単位としてまとめた上で、単純無作為抽出で世帯を抽出する方法により行いつつ、個人を単位とするレコードでリサンプリング率が約80%になるようにしているが、適切か。

エ 情報の削除

i) 直接的な識別情報の削除

調査地域を特定する調査区番号などの実査用の識別番号や調査客体を直接識別できる情報は削除することとしているが、適切か。

ii) 出現頻度が低い又は特徴的な値があるレコードを含む世帯の削除

出現頻度が低い又は特徴的な値がある以下のレコードを含む世帯を削除することとしているが、適切か。

①世帯人員が多い世帯

世帯人員数9人以上の世帯を削除

②子供の数の多い世帯

国勢調査（母集団情報）において、親の年齢、子供の数及び住宅の所有の関係を組み合わせた結果、発生頻度の低い組合せに該当する世帯を削除

③母子世帯及び父子世帯

国勢調査（母集団情報）において、母又は父の年齢、子供の数及び住宅の所有の関係を組み合わせた結果、発生頻度の低い組合せに該当する世帯を削除

- ④三つ子以上がいる世帯
三つ子以上がいる世帯を削除

オ 分類区分の再編（世帯員に関する項目）

i) 年齢

世帯員の年齢については、0～9歳を各歳、10～84歳を5歳階級区分としているが、適当か。

また、85歳以上をトップコーディングすることは適当か。

ii) 末子年齢

末子の年齢については、0歳、1～2歳、3～5歳、6～8歳、9～11歳、12歳以上としているが、適当か。

また、12歳以上をトップコーディングすることは適当か。

(2) 匿名化措置をしていない事項

匿名化措置をしていない事項のうち、外観的に容易に特定化されるおそれがあるため匿名化の措置をすべき事項はないのか。

2 前回答申における「今後の課題」への対応

前回の答申^(注)において「今後の課題」とされた以下の事項への対応は適当か。

(1) 複数の匿名データの作成

前回の答申における「匿名化措置を課す情報及びその程度が異なる複数の匿名データの作成の可能性について検討する」についての対応は適当か。

(2) 匿名データの提供時期の短縮化

前回の答申における「調査実施後5年以上経過したものを提供するという基準を緩和することについて検討する」についての対応は適当か。

(3) トップコーディング等が行われた変数

前回の答申における「トップコーディング等が行われた変数についても多変量解析に十分利用できるよう、当該変数の平均値等をメタデータとして整備する」についての対応は適当か。

(注) 諮問第13号の答申は、社会生活基本調査のほか、全国消費実態調査、就業構造基本調査及び住宅・土地統計調査を含めた統計調査の匿名データ作成に係る答申