

「国勢調査に係る匿名データの作成について」の論点

平成 24 年 12 月 7 日

椿 広 計

1 匿名データの作成方法について

(1) 匿名化措置を予定している事項

「平成 12 年国勢調査（大規模調査）」及び「平成 17 年国勢調査（簡易調査）」において、匿名化措置を予定している次の事項は、国勢調査が悉皆調査であることや、詳細な集計表が外部参照情報として利用可能であることなどの特性も踏まえて、匿名性の確保の観点から確実なものとなっているか。

一方、有用性の確保の観点から、利用者にとって一定程度の利用ができるものとなっているか。過剰な匿名化措置により、多くの利用者が利用しにくいものとなっているか。

ア 作成に使用するデータ

調査票に記入された内容を数値や符号に置き換えた全数データの電磁的記録を用いる計画としているが、適当か。

イ 地域区分

地域区分は、「都道府県」及び「人口 50 万以上の市区」とする計画であるが、匿名性及び有用性の観点から適当か。

ウ サンプリングの方法等

全世帯を母集団として 1%をサンプリングすることとし、世帯の種類（一般世帯及び施設等の世帯）ごとに、市区町村及び世帯人員等で並べ替えた上で、「一般世帯」については世帯単位で、また「施設等の世帯」については個人単位で、それぞれ系統抽出を行い、これらを統合して匿名データとする計画であるが、匿名性の観点から、抽出率及びサンプリングの方法は適当か。「一般世帯」と「施設等の世帯」で抽出単位が異なっているが、ウエイトなどの必要性を考慮すべきか。

また、公表統計との間で、主要な項目の分布などに大きな乖離はなく、当該データの有用性が確保されているか。

エ 情報の削除

(ア) 直接的な識別情報の削除

人口 50 万未満の市区町村コード、調査区番号、世帯番号・調査区内連番など、調査客体を直接識別できる情報を削除することとしているが、適当か。

(イ) 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、適当か。

① 世帯人員が多い世帯

地域区分ごとの出現頻度により、世帯人員が 7 人以上～9 人以上いる世帯を削除

②父子世帯

未婚、死別又は離別の男親と、その未婚の20歳未満の子供のみからなる世帯を削除

③年齢差の大きい夫婦のいる世帯

年齢差が25歳以上の夫婦のいる世帯を削除

④年齢差の大きい又は小さい親子のいる世帯

年齢差が45歳以上の親と子、年齢差が14歳以下の親と長子又は19歳以下の親と末子のいる世帯を削除

⑤世帯主又は配偶者のいずれか一方若しくは双方が外国人で子供の数が多い世帯

地域区分ごとの出現頻度により、子供の数が3人以上～7人以上の外国人世帯を削除

(ウ) 既存の統計表により母集団一意又は二意であることが判明しているレコードを含む世帯の削除

既存の統計表により、全国において、母集団一意又は二意であることが判明しているレコードが含まれる世帯を削除することとしているが、適当か。

加えて、既存の統計表のうち、特に外観識別性が高いと考えられる項目が含まれる統計表については、地域（都道府県、人口50万人以上の市区）において、母集団一意又は二意であることが判明しているレコードが含まれる世帯も削除することとしているが、適当か。

オ 世帯員に関する項目の再編等

(ア) 年齢

世帯員の年齢については、0～84歳を5歳階級区分とする計画である。各歳別のデータ提供に対するニーズは大きいと考えられるが、匿名性及び有用性の観点から適当か。

また、85歳以上をトップコーディングすることは適当か。年々高齢化が進展している中で、地域によっては85歳以上の割合が高くなってきていると考えられるが、問題はないか。

(イ) 世帯主との続き柄

「他の親族」、「住み込みの雇人」及び「その他」を統合することは適当か。

(ウ) 国籍

「日本人」及び「外国人」の2区分とし、外国籍の内訳は提供しない計画であるが、適当か。

(エ) 5年前の住居の所在地（平成12年国勢調査）

「他県から」及び「国外から」を統合する計画（5年前の常住都道府県等は提供されない。）であるが、匿名性及び有用性の観点から適当か。

(オ) 労働力状態

「家事などのほか仕事」及び「通学のかたわら仕事」を統合することは適当か。

(カ) 就業時間

実数、及び既存の統計表と同じ階級（14区分及び10区分）で提供することとし、実数については、90時間以上をトップコーディングする計画であるが適当か。

(キ) 従業上の地位

「雇人のある業主」、「雇人のない業主」及び「家庭内職者」を統合することは適当か。

(ク) 産業（大分類）

産業については、大分類で提供し、以下の分類区分の統合を行う計画であるが、匿名性及び有用性の観点から適当か。産業の性質や従事者の属性をより考慮すべきか。

- ①「農業」、「林業」及び「漁業」を統合
- ②「鉱業」及び「建設業」を統合
- ③「製造業」及び「電気・ガス・熱供給・水道業」を統合
- ④「複合サービス業」及び「サービス業（他に分類されないもの）」を統合（平成17年国勢調査）

(ケ) 職業（大分類）

職業については、大分類で提供し、「保安職業従事者」、「農林漁業作業者」及び「運輸・通信従事者」を統合する計画であるが、匿名性及び有用性の観点から適当か。職業の性質や従事者の属性をより考慮すべきか。

(コ) 常住地による従業地・通学地

「県内他市区町村で従業・通学」及び「他県で従業・通学」を統合することは適当か。

(サ) 利用交通手段（平成12年国勢調査）

「利用交通手段が1種類」について、「勤め先・学校のバス」、「ハイヤー・タクシー」、「オートバイ」及び「その他」を統合することは、利用交通手段の把握として適当か。

カ 世帯に関する項目の再編等

(ア) 世帯の種類

「一般世帯」及び「施設等の世帯」の2区分を提供することとし、「施設等の世帯」の内訳は提供しない計画であるが適当か。

(イ) 世帯人員

「施設等の世帯」については提供しない計画であるが適当か。

(ウ) 世帯の家族類型

6区分による提供を計画しているが適当か。

(エ) 家計の収入の種類（平成12年国勢調査）

「家計の収入の種類」は提供しない計画であるが、匿名性及び有用性の観点から適当か。

(オ) 住居の種類・住宅の所有の関係

「住宅に住む一般世帯」について、以下の分類区分を統合する計画であるが適当か。

- ①「公営の借家」及び「都市機構・公社の借家」を統合
- ②「給与住宅」及び「間借り」を統合

また、「住宅以外に住む一般世帯」の内訳（「会社等の独身寮」及び「その他」）は提供しない計画であるが適当か。

(カ) 住宅の床面積

実数ではなく、基本的に既存の統計表に合わせた階級で提供することとし、この際、「200～249㎡」及び「250㎡以上」を統合する計画であるが適当か。

(キ) 住宅の建て方

「長屋建」及び「その他」を統合することは適当か。

(ク) 建物全体の階数

実数ではなく、基本的に既存の統計表に合わせた階級で提供することとし、地域区分ごとの出現頻度により、必要に応じて、「6～10階建」、「11～14階建」、「15階建以上」の区分のうち、「6～10階建」以上、あるいは「11～14階建」以上の区分を統合する計画であるが適当か。

(ケ) 世帯が住んでいる階

実数ではなく、基本的に既存の統計表に合わせた階級で提供することとし、地域区分ごとの出現頻度により、必要に応じて、「3～5階」、「6～10階」、「11～14階」、「15階以上」の区分のうち、「3～5階」以上、「6～10階」以上、あるいは「11～14階」以上の区分を統合する計画であるが適当か。

キ スワッピング

今回、地域区分として「都道府県」及び「人口50万以上の市区」の情報を提供することとされているものの、地域情報は強力な識別情報となり得ることから、一部世帯を他の地域の類似世帯と入れ替える計画であるが、匿名性の観点からスワッピングの導入及びその方法は適当か。また、スワッピングの前後で影響を受ける項目の分布にあまり差異がないなど、有用性が確保されているか。

(2) 匿名化措置を予定していない事項

匿名化措置が予定されていないが、外観的に容易に把握されるおそれがある等により、匿名化を確保すべき事項は存在しないか。

2 他の情報との関係について

外部の情報との対応関係から世帯・個人を特定される危険性はないか。

3 その他

(1) トップコーディングが行われた変数の基本統計量の提供

トップコーディングが行われた変数については、利用者の利便性向上の観点から当該変数の基本統計量を提供することはできないか。

(2) 匿名データの提供時期の短縮化

今回、匿名データの提供時期を調査実施後5年以上経過したものから提供することとしているが、一方で、直近の統計による分析の重要性も増していることから、匿名性及び有用性の観点から適当か。

以上