

諮問第 108 号の答申

住宅・土地統計調査に係る匿名データの作成について

本委員会は、諮問第 108 号による住宅・土地統計調査に係る匿名データの作成について審議した結果、下記のとおり結論を得たので、答申する。

記

1 計画の適否

平成 29 年 10 月 26 日付け諮問第 108 号「住宅・土地統計調査に係る匿名データの作成について」(以下「本計画」という。)を審議した結果、以下のとおり、調査回答者の匿名性及び学術研究等における有用性が確保されるものと認められることから、本計画で平成 20 年及び平成 25 年の住宅・土地統計調査の匿名データを作成することは、適当である。

2 理由等

(1) 上限値、下限値の基準値の設定措置

本計画では、「建物の建築面積」や「1か月当たり家賃・間代」などの項目について、一定の値を上限値とし、それを上回る場合に上限値以上でまとめる措置(以下「トップコーディング」という。)、また、一定の値を下限値とし、それを下回る場合に下限値以下でまとめる措置(ボトムコーディング)をとっている。これまで、この基準値を全国一律で設定していたが、本計画の対象調査年次である平成 20 年及び平成 25 年の住宅・土地統計調査の匿名データ作成では、都道府県別に基準値を設定することにしている。

これについては、建築面積、家賃など都道府県によって分布が異なる項目の情報がより詳細に提供され有用性が高まることから、適当である。

(2) 新規追加調査項目の匿名化措置

本計画では、平成 20 年調査において新規追加された「住宅の耐震診断の有無」及び「オートロックの別」の調査項目は、各回答欄の回答数が一定数存在するため、そのまま提供することになっている。

これについては、匿名性が確保され有用性が高まることから、適当である。

一方、平成 25 年調査における新規追加調査項目のうち「東日本大震災による転居」及び継続調査項目のうち「住宅の増改築 改修工事等」に追加された設問項目「東日本大震災による被災箇所の改修工事をした」については、対象となる世帯が限られており、当該項目を複数の項目と組み合わせることにより、世帯が特定されるリスクが高まるため、当該項目は提供しないことにしている。

これについては、当該項目が、東日本大震災の影響を分析できる有用な項目である一方、匿名データとして作成する場合、匿名性の確保に支障を来すおそれがあることから、提供しないことはやむを得ない。

(3) 諮問第 13 号の答申(平成 21 年 3 月 9 日答申、府統委第 22 号)における「今後の課題」への対応

住宅・土地統計調査の匿名データ作成については、諮問第 13 号の答申時において、①複数の匿名データ作成の可能性、②提供する調査年次の拡張、提供開始までの期間短縮、③トップコーディング等が行われた変数の平均値等の整備が課題として指摘されていた。

ア 複数の匿名データ作成の可能性

既存の匿名データとは異なり、都道府県より詳細な地域情報などを提供する新たな匿名データを作成することについて、本計画では、匿名データの有用性と個人情報の秘匿性確保の双方を考慮した作成方法など基礎的な課題が多いことから、平成 22 年国勢調査の匿名データ作成の検討の中で分析を進め、その検討結果も踏まえた上で、検討することとしている。

これについては、統計委員会の意向を踏まえ総務省において昨年実施した匿名データに地域情報を付与する調査研究においても、地域情報の詳細化のためには十分な検討が必要であるとの方向が示されていることから、現段階で結論が得られないことはやむを得ない。なお、総務省統計局においては、国勢調査の匿名データ作成の検討と併せて早期の検討が望まれる。

イ 提供する調査年次の拡張、提供開始までの期間短縮

提供する調査年次の拡張に関しては、平成元年以前のデータと直近のデータの 2 種類が対象であるが、本計画では、よりニーズの高い直近の年次データである平成 20 年及び 25 年調査の提供を優先している。

これについては、直近データを優先して提供することが利用者ニーズに対応するものであり、適当である。ただし、平成 20 年調査は、提供時期としては遅いため、答申後、速やかに作成・提供することに努めるべきである。

また、提供開始までの期間短縮について、本計画では、調査実施後 5 年以上経過したものを提供していたことも含め、今後、検討することとしている。

これについては、リソースの制約がある中で、匿名データの有用性と個人情報の秘匿性確保、双方の視点を踏まえつつ検討する必要があることからやむを得ない。今後、引き続き提供を早期化することが望まれる。

ウ トップコーディング等が行われた変数の平均値等の整備

本計画では、多変量解析等に資するため、トップコーディング等が行われた変数について、その統合された部分に関する平均と標準偏差を都道府県別に提供することとしている。

これについては、有用性がより高まることから、適当である。