

「全国消費実態調査等に係る匿名データの作成について」の論点

平成 21 年 2 月 13 日

廣 松 毅

1 匿名化措置の予定事項

匿名化措置を予定している下記の事項は、匿名性の確保の観点から確実なものとなっているか。調査対象者が特定されるリスクが十分低いものとなっているか。

一方、有用性の確保の観点から、利用者が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ないか。

(1) 4 調査共通事項**ア リサンプリングの方法及び地域区分**

リサンプリングの方法や再付与する乗率の算出方法は適切か。

リサンプリングの方法として、世帯ごとにサブサンプルを抽出すると、1 ファイル内に世帯員全員に係る情報が含まれ調査客体が特定し易くなることから、個人単位でサブサンプルを抽出する方法を検討する必要はないか。

結果精度を向上させる目的で人口の少ない地域について標本抽出率を高く設定している調査の場合、当該地域における特定リスクの防止の観点から、標本設計の内容を踏まえてリサンプリング方法を検討する必要はないか。

「全国消費実態調査」等 3 調査については、80%のリサンプリング率を前提として地域区分を 6 ブロックとしているが、このリサンプリング率と地域区分の組み合わせは適切か。データが特定の都道府県に偏在しており、地域区分の効果を喪失させている例はないか。

「住宅・土地統計調査」については 10%のリサンプリング率を前提として地域区分を 47 都道府県としているが、このリサンプリング率と地域区分の組み合わせは適切か。人口が著しく少ない県の場合でも 10%のリサンプリング率で問題がないか。

リサンプリングデータは、公表されている基データによる統計との整合性等から見て、データ利用上、有効なものか。

イ 裾きり及びトップ（ボトム）コーディングの基準

全体に占める構成比が 0.5%未満と極めて小さい変数はレコード削除又はコーディングを行うとの考え方は、該当する全ての変数に適用すべきではないか。

ごく一部の都道府県のデータが構成比 0.5%未満である場合、全ての都道府県のデータを一つ上（下）の階級に統合してコーディングするのではなく、該当する都道府県のデータのみコーディングすれば良いのではないか。

世帯人員が 8 人以上の世帯のレコードを削除することは適切か。世帯人員が 8 人以上の世帯は、全体に占める構成比が全国的に見ると小さくても地域別に見ると必ずしも小さくない場合もあるため、当該削除により有用性が損なわれるおそれがあるのではないか。また、8 人以上という区切りは適切か。

トップコーディングをする年齢（全国消費実態調査及び住宅・土地統計調査は 75 歳以上、社会生活基本調査は 85 歳以上、就業構造基本調査は 80 歳以上）は適切か。

トップ（ボトム）コーディングされた変数は、集計値の分析上には問題ないが、回帰分

析には利用できないため、当該変数の平均値等をメタデータとして利用者に提供する等の措置を講じる必要はないか。

ウ 世帯員の年齢のリコーディング

原則として5歳階級別リコーディングすることは適当か。就業構造基本調査については、就業行動の分析への十分な活用の観点から、主要な世帯員については、各歳別に提供すべきではないか。

15歳未満の世帯員について、各歳別に提供することは適当か。
(リコーディングされた変数について上記イと同じ)

(2)全国消費実態調査

ア 地理的情報

標本データを母集団に復元するための乗率と地理的情報を組み合わせることで市町村レベルまで特定されることを防ぐため、リサンプル時に乗率を再付与する措置は適当か。

イ 世帯の住居

トップ(ボトム)コーディングする住居の延べ床面積、住宅の敷地面積等(延べ床面積200㎡以上はトップコーディング・二人以上世帯の30㎡未満はボトムコーディング、住宅の敷地面積1,000㎡以上をトップコーディング等)は適当か。

ウ 世帯主等の年収・世帯の貯蓄負債

世帯主等の年収等が一定金額以上の世帯(世帯主等の年収6,000万円以上、世帯の貯蓄現在高3億円以上、月賦・年賦の未払い残高1億円以上、住宅・土地の購入等の借入金残高3億円以上)を削除することは適当か。また、例えば、貯蓄現在高が3億円未満であれば、定期貯金等が稀な金額であっても匿名化措置を講じる必要はないのか。

(3)就業構造基本調査

ア 世帯主の年齢、夫の年齢、妻の年齢、離職時の年齢

世帯主の年齢、夫の年齢、妻の年齢及び離職時の年齢について、5歳階級グルーピングを施すことは適当か。

イ 特定世帯の親の年齢

特定世帯(母子世帯・父子世帯等)の男(女)親の年齢について、5歳階級グルーピングを施すことは適当か。近年80歳以上の親が急激に増加していることを踏まえ、親子関係の分析の観点から、80歳代も5歳階級別に提供し、トップコーディングの上限値を90歳以上に引き上げるべきではないか。

(4)住宅・土地統計調査

ア 地理的情報(全国消費実態調査と同じ)

イ 世帯の住居(戸建て)

延べ(床)面積(全国消費実態調査と同じ)

居住室数、居住室の畳数

居住室数については10室以上、居住室の畳数については60畳以上をトップコーディングすることとしているが、この区切りは適当か。当該区切りを引き下げる必要はないか。また、居住室の畳数について9畳未満をボトムコーディングすることとしているが、

この区切りは適当か。

敷地面積、建築面積

敷地面積 700 m²以上をトップコーディング、50 m²未満をボトムコーディングすることとしているが、この区切りは適当か。

建築面積 150 m²以上をトップコーディング、30 m²未満をボトムコーディングすることとしているが、この区切りは適当か。

家賃・間代

家賃等について9万円以上でトップコーディングすることとしているが、この区切りは適当か。各都道府県の上限值を揃える必然性は少ないのではないか。

ウ 世帯の住居（共同住宅等）

敷地面積

敷地面積 2,000 m²以上をトップコーディング、100 m²未満をボトムコーディングすることとしているが、この区切りは適当か。

建築面積

建築面積 500 m²以上をトップコーディング、100 m²未満をボトムコーディングすることとしているが、この区切りは適当か。

階数

共同住宅については、11階以上をトップコーディング、2階未満をボトムコーディング、長屋建については、2階以上をトップコーディングすることとしているが、この区切りは適当か。

エ 外部の情報

調査客体を特定できる可能性のある外部情報として不動産登記簿があるものの、住居に関する事項はトップ（ボトム）コーディングを施していることから、当該登記簿との照合は事実上困難としているが、この判断は適当か。

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、匿名性の確保の観点から当該措置の必要がある事項はないか。例えば、従事者数の少ない職業への就業者、稀有の特徴のある住宅など外観的に容易に把握されるおそれがあるものはないか。

3 その他

1調査当たり匿名化措置を施した1種類の匿名データのみ作成することとしているが、利用者の種類や利用の用途等に応じて、匿名化措置のレベルを変えた複数の匿名データ（一般研究者用、訓練・教育用等）を作成する必要はないか。

作成対象年を、直近の調査や昭和以前に実施された調査に拡大する必要はないか。