

「国民生活基礎調査に係る匿名データの作成について」の論点

平成 23 年 2 月 7 日

椿 広 計

1 匿名性の確保と有用性の確保**(1) 匿名化措置を予定している事項**

平成 16 年国民生活基礎調査（以下「本調査」という。）において、匿名化措置を予定している次の事項は、匿名性の確保の観点から確実なものとなっているか。既に提供されている全国消費実態調査等の 4 つの調査（以下「統計局 4 調査」という。）とは異なる標本抽出の方法（集落抽出）を採用しているが、このことによる開示リスクは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ないか。

ア リサンプリングの単位及び割合等

(ア) 世帯票及び健康票から作成される匿名データ（以下「匿名データ A」という。）は、二段抽出を行った結果として約 2 割のリサンプリング率となっているが、匿名性の観点からリサンプリング率及び二段抽出の方法は適当か。また、公表統計との間で、代表的な項目の平均値・質的項目の比率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。

(イ) 世帯票、健康票、所得票及び貯蓄票から作成される匿名データ（以下「匿名データ B」という。）については、二段抽出を行った結果として約 2 割のリサンプリング率となっているが、匿名性の観点からリサンプリング率及び二段抽出の方法は適当か。また、公表統計との間で、世帯の所得や貯蓄の分析をする際に代表的な項目の平均値・度数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。

イ 地域区分

本調査の集落抽出であることによる開示リスクを十分低減するため、地域区分を「全国」一本とする予定であるが、有用性の観点から、何らかの地域表章が可能かどうかについて、リサンプリング率など、他の秘匿措置と併せて検討すべきではないか。

ウ 世帯人員

世帯人員が 8 人以上の世帯のレコードは削除することとしているが、この区切りは適当か。

エ 世帯員の年齢

(ア) 15 歳以上 85 歳未満の世帯員について、5 歳階級別にグルーピングすることは適当か。また、15 歳未満の世帯員について、0～5 歳、6～11 歳及び 12～14 歳の 3 区分とすることとしているが、匿名性の確保の観点から適当か。また、統計局 4 調査のように各歳別の提供はできないか。

(イ) 年齢 85 歳以上の世帯員について、トップコーディングをすることは適当か。

オ 同一年齢の子供が 3 人以上いる世帯、父子世帯及び要介護者が 2 名以上いる世帯

3 つ子以上世帯、父子世帯及び要介護者 2 名以上世帯をレコード削除することは適当か。

カ その他の秘匿措置

- (ア) トップ（ボトム）コーディングをかける場合、本調査では対象全体の1%未満対象としているが、これは秘匿性及び有用性の観点から適当か。（統計局4調査の場合、0.5%未満をトップ（ボトム）コーディングしているため。）
- (イ) 匿名データBのうち、世帯票及び健康票に係る特定の量的項目に関しては、秘匿性及び有用性の観点から、所得票及び貯蓄票の調査対象のみの分布に基づいてトップ（ボトム）コーディングを適用する方が望ましいのではないか。
- (ウ) 地域区分を「全国」のみ提供することとしていることから、例えば「1週間に仕事をした時間」のトップコーディングなど、外観から識別される可能性が低いと思われる項目の秘匿措置に関して、緩和することはできないか。また、他にも秘匿措置を緩和できる項目はないか。
- (エ) 安全性及び有用性の観点から、トップ（ボトム）コーディングした階級の平均値などについて、併せて提供することはできないか。

キ 所得票及び貯蓄票に関する事項

世帯票及び健康票のデータも含めると、豊富な世帯情報が提供されることとなるが、特定される危険性は十分に低いものとなっているか。また、諮問第13号の答申及び当該諮問に係る匿名データ部会の議論を踏まえ、所得票及び貯蓄票では所得や貯蓄の世帯総額のみ公開し、その内訳や世帯員別の金額についてはすべて削除することとしているが、所得や貯蓄の分析の観点からさらに提供できる項目はないか。

(2) 匿名化措置を予定していない事項

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。一方、有用性の観点から提供が望まれるデータはないか。

ア 外観識別可能な事項

同一年齢の子供が3人以上いる世帯、父子世帯及び要介護者が2名以上いる世帯以外に、外観識別可能な世帯員構成等は存在しないか。

イ 介護票の匿名データ化

介護票はデータ数が少なく、リサンプリングすることにより有用性の確保が困難であること及び介護の状況は外観から識別できることを踏まえ、介護票のデータからは匿名データの作成を行わないこととしているが、有用性の観点からスワッピングや誤差の付加などといった上記措置以外の匿名化技法を措置することで匿名データの作成はできないか。

ウ 世帯員単位での匿名データ化

今回、世帯員単位での匿名データの作成は行われていないが、公衆衛生・疫学分野等における有用性の観点から、世帯員単位での匿名データの作成はできないか。

2 他の情報との関係

(1) 後続調査との関係

本調査を親標本とした後続調査として、国民健康・栄養調査、世帯動向調査、公的年金加入状況等調査等、厚生労働省所管の様々な一般統計調査が実施されているが、本調査の匿名データの提供によりこれら後続調査の調査客体が特定される危険性は十分低いものとなっているか。

(2) 統計調査以外の情報との関係

ア 外部情報との照合可能性

本調査の調査設計が集落抽出によることから、一般的な標本抽出に比べ開示リスクが高いと思われるが、地方公共団体等が保有する外部情報との対応関係から特定される危険性は十分に低いものとなっているか。

イ 照合の対象とすべき外部情報の基準

一般に行政記録情報を開示する場合、他の外部情報との照合による識別も含めて、特定の個人等が識別できないようにするが、匿名データの提供においては、いわゆる一般人が入手し得る情報での照合が可能な場合（一般人基準）と、特定人が持っている情報との関係で照合が可能な場合（特定人基準）のいずれの基準を照合の対象となる他の外部情報とするのか。

3 その他

(1) 作成対象年次等の拡大

今回、平成 16 年調査のみを匿名データの作成対象としているが、経年的な分析の観点からは、過去のデータについても匿名データが提供されるべきではないか。

また、本調査は、3 年ごとの大規模調査とそれ以外の年に実施する中間年調査で構成されているが、経年的な分析の観点から、過去のデータのみならず、中間年調査についても匿名データの作成を検討する必要はないのか。

(2) 提供時期の変更

今回、統計局 4 調査の対応及び本調査の特性を踏まえて、匿名データの提供時期を調査実施後 5 年経過してから提供することとしているが、さらに短い期間での提供をすることはできないか。

(3) 当該統計調査への影響

行政記録情報を開示する際、他者からは十分に匿名化されていても本人は識別できる場合に、そのことが事務・事業に支障を及ぼす恐れがある場合には開示しないとするところもあるが、統計調査の場合、調査への協力に悪影響を及ぼす恐れについて考慮する必要はないか。

(見え消し版)

「国民生活基礎調査に係る匿名データの作成について」の論点(案)

平成 2223 年 422 月 247 日
椿 広 計

1 匿名性の確保と有用性の確保

(1) 匿名化措置を予定している事項

平成 16 年国民生活基礎調査(以下「本調査」という。)において、匿名化措置を予定している次の事項は、匿名性の確保の観点から確実なものとなっているか。既に公開提供されている全国消費実態調査等の 4 つの調査(以下「統計局 4 調査」という。)とは異なる標本抽出の方法(集落抽出)を採用しているが、このことによる開示リスクは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ないか。

ア リサンプリングの単位及び割合等

—(7) 世帯票及び健康票から作成される匿名データ(以下「匿名データ A」という。)は、二段抽出によりを行った結果として 20%を約 2 割のリサンプリング率となっずるとして、匿名性の観点からリサンプリング率及び二段抽出の方法は適当か。また、公表統計との間で、代表的な項目の平均値・質的項目の比率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。

—(4) 世帯票、健康票、所得票及び貯蓄票から作成される匿名データ(以下「匿名データ B」という。)については、二段抽出によりを行った結果として 20%を約 2 割のリサンプリング率となっずるとして、匿名性の観点からリサンプリング率及び二段抽出の方法は適当か。また、公表統計との間で、世帯の所得や貯蓄の分析をする際に代表的な項目の平均値・度数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。

イ 地域区分

本調査の集落抽出であることによる開示リスクを十分低減するため、地域区分を「全国」一本とする予定であるが、匿名性の確保の観点から適当か。一方、有用性の観点から、何らかの地域表章~~は~~が可能かどうかについて、リサンプリング率など、他の秘匿措置と併せて検討すべきではないか。

ウ 世帯人員

世帯人員が 8 人以上の世帯のレコードは削除することとしているが、この区切りは適当か。

エ 世帯員の年齢

—(7) 15 歳以上 85 歳未満の世帯員について、5 歳階級別にグルーピングすることは適当か。また、15 歳未満の世帯員について、0～5 歳、6～11 歳及び 12～14 歳の 3 区分とすることとしているが、匿名性の確保の観点から適当か。また、統計局 4 調査のように各歳別の提供はできないか。

—(4) 年齢 85 歳以上の世帯員について、トップコーディングをすることは適当か。

オ 同一年齢の子供が 3 人以上いる世帯、父子世帯及び要介護者が 2 名以上いる世帯

(見え消し版)

3つ子以上世帯、父子世帯及び要介護者2名以上世帯をレコード削除することは適当か。

カ その他の秘匿措置

(ア) トップ(ボトム)コーディングをかける場合、本調査では対象全体の1%未満対象としているが、これは秘匿性及び有用性の観点から適当か。(統計局4調査の場合、0.5%未満をトップ(ボトム)コーディングしているため。)

(イ) 匿名データBのうち、世帯票及び健康票に係る特定の量的項目に関しては、秘匿性及び有用性の観点から、所得票及び貯蓄票の調査対象のみの分布に基づいてトップ(ボトム)コーディングを適用する方が望ましいのではないか。

(ウ) 地域区分を「全国」のみ提供することとしていることから、例えば「1週間に仕事をした時間」のトップコーディングなど、外観から識別される可能性が低いと思われる項目の秘匿措置に関して、緩和することはできないか。また、他にも秘匿措置を緩和できる項目はないか。

(エ) 安全性及び有用性の観点から、トップ(ボトム)コーディングした階級の平均値などについて、併せて提供することはできないか。

カキ 所得票及び貯蓄票に関する事項

世帯票及び健康票のデータも含めると、豊富な世帯情報が公開提供されることとなるが、特定される危険性は十分に低いものとなっているか。また、諮問第13号の答申及び当該諮問に係る匿名データ部会の議論を踏まえ、所得票及び貯蓄票では所得や貯蓄の世帯総額のみ公開し、その内訳や世帯員別の金額についてはすべて削除することとしているが、所得や貯蓄の分析の観点からさらに公開提供できる項目はないか。

(2) 匿名化措置を予定していない事項

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。一方、有用性の観点から提供が望まれるデータはないか。

ア 外観識別可能な事項

同一年齢の子供が3人以上いる世帯、父子世帯及び要介護者が2名以上いる世帯以外に、外観識別可能な世帯員構成等は存在しないか。

イ 介護票の匿名データ化

介護票はデータ数が少なく、リサンプリングすることにより有用性の確保が困難であること及び介護の状況は外観から識別できることを踏まえ、介護票のデータからは匿名データの作成を行わないこととしているが、有用性の観点からスワッピングや誤差の付加などといった上記措置以外の匿名化技法を措置することで匿名データの作成はできないか。

ウ 世帯員単位での匿名データ化

今回、世帯員単位での匿名データの作成は行われていないが、公衆衛生・疫学分野等における有用性の観点から、世帯員単位での匿名データの作成はできないか。

2 他の情報との関係

(1) 後続調査との関係

本調査を親標本とした後続調査として、国民健康・栄養調査、世帯動向調査、公的年金加入状況等調査等、厚生労働省所管の様々な一般統計調査が実施されているが、本調査の匿名データの公開が提供によりこれら後続調査の調査客体が特定される危険性は十分に低いものと

(見え消し版)

なっているか。

(2) 統計調査以外の情報との関係

ア 外部情報との照合可能性

本調査の調査設計が集落抽出によることから、一般的な標本抽出に比べ開示リスクが高いと思われるが、地方公共団体等が保有する外部情報との対応関係から特定される危険性は十分に低いものとなっているか。

イ 照合の対象とすべき外部情報の基準

一般に行政記録情報を開示する場合、他の外部情報との照合による識別も含めて、特定の個人等が識別できないようにするが、匿名データの提供においては、いわゆる一般人が入手し得る情報での照合が可能な場合（一般人基準）と、特定人が持っている情報との関係で照合が可能な場合（特定人基準）のいずれの基準を照合の対象となる他の外部情報とするのか。

3 その他

(1) 作成対象年次等の拡大

今回、平成 16 年調査のみを匿名データの作成対象としているが、経年的な分析の観点からは、過去のデータについても匿名データが提供されるべきではないか。

また、本調査は、3 年ごとの大規模調査とそれ以外の年に実施する中間年調査で構成されているが、経年的な分析の観点から、過去のデータのみならず、中間年調査についても匿名データの作成を検討する必要はないのか。

(2) 提供時期の変更

今回、統計局 4 調査の対応及び本調査の特性を踏まえて、匿名データの提供時期を調査実施後 5 年経過してから提供することとしているが、さらに短い期間での提供をすることはできないか。

(3) 当該統計調査への影響

行政記録情報を開示する際、他者からは十分に匿名化されていても本人は識別できる場合に、そのことが事務・事業に支障を及ぼす恐れがある場合には開示しないとするところもあるが、統計調査の場合、調査への協力に悪影響を及ぼす恐れについて考慮する必要はないか。