

諮問第 37 号の答申 労働力調査に係る匿名データの作成について（案）

本委員会は、総務省が作成を予定している労働力調査（以下「本調査」という。）に係る匿名データの作成方法の計画について審議した結果、下記の結論を得たので答申する。

記

1 計画の適否

本計画については、これにより作成される匿名データにおいて、本調査の調査客体の匿名性及び学術研究等における有用性がおおむね確保されるものと認められることから、適当である。

2 理由等

(1) 情報の削除

ア レコードのリサンプリング

本調査の匿名データの作成に当たっては、地域 11 ブロック及び組符号 8 区分による層化を行ったのち、世帯単位により、約 80%（沖縄県については約 20%）を等確率で再抽出（以下「リサンプリング」という。）したもの（以下「サブサンプル」という。）を用いる計画である。

これらについては、次の理由等から適当である。

- ① リサンプリングは、匿名データの中に特定の調査客体が含まれるか否かの判断を困難とする措置であること
- ② 本調査は、層化 2 段抽出法により地域 11 ブロック別に調査客体を選定するとともに、標本の交代を 8 組のグループに分けて行っていることから、地域 11 ブロック及び組符号 8 区分による層化を行ったのち、約 80% を等確率でリサンプリングすることにより、抽出結果の安定性を図っていること
- ③ 沖縄県については、集計乗率が他都道府県と比較して小さいことから、リサンプリング率を 20% とした上で各レコードに 4 倍の乗率を再付与することにより、集計乗率から地域を特定されることを防ぐための措置を取っていること
- ④ 世帯単位による抽出は、利用者のニーズの高い、世帯構成に着目した世帯員の就業・不就業の状況等に関する分析が可能となるため、世帯員単位による抽出よりも当該データの有用性が高まること
- ⑤ 当該方法によりリサンプリングされたサブサンプルから作成された匿名データによる統計と全レコードから作成された公表統計（以下「公表統計」という。）との間で、労働力人口比率や完全失業率といった代表的な比率の値に大きな乖離は無く、当該データの有用性が確保されていること

イ 識別情報の削除等

(ア) 地理的情報の削除等

本調査のサブサンプル中のレコードに含まれる情報のうち、調査区符号、標本符号、世帯符号を削除し、地域区分を全国 1 区分とするとともに、レコードは世帯単位にランダムに並

べ替えを行う計画である。

このうち、調査区符号や世帯符号等を削除し、レコードをランダムに並べ替えることについては、調査客体の特定や探索を防止するために効果的な措置であることから、適当である。なお、本調査では、同一の住戸に居住する世帯に対して2か月連続して調査が行われ、翌年の同期にさらに2か月継続して調査が行われることから、世帯毎に異時点間の情報をつなぎ合わせる事が可能となるような情報の提供に対するニーズも指摘されているが、異動情報の組合せによっては出現頻度が極めて低く、調査客体を特定される可能性が生じることから、当該情報を提供しないことは適当である。

また、地域区分を全国1区分とすることについては、本計画におけるサブサンプルの抽出率が約80%と高く、「事業の種類（産業）」や「本人の仕事の種類（職業）」等の個人の仕事の内容を示す外観識別可能性の高い属性情報と詳細な地域情報を組み合わせた場合に、調査客体を特定される可能性が生じる事等から、やむを得ない措置である。

(イ) 前月欄の情報の削除

本調査は、集計上、2か月目のデータには前月の情報が一部保持されているが、この前月欄に含まれる情報のうち、「従業上の地位」、「事業の種類（産業）」、「勤め先・業主などの企業全体の従業者数」については、これを削除し、「月末1週間に仕事をしたかどうかの別」のみを提供する計画である。

「月末1週間に仕事をしたかどうかの別」を提供することについては、前月からの異動に関する情報は、就業状態の遷移分析において非常に重要な情報であること、当該異動情報のみから調査客体を特定される可能性は低いことから、適当である。

一方、「事業の種類（産業）」等の情報を削除することについては、これらが外部から比較的容易に把握可能な属性情報であり、例えば、産業分類を大括り化したとしても、異動情報を組み合わせることによって調査客体を特定される可能性が生じることから、適当である。

ウ 裾切りによるレコード削除

(ア) 世帯人員が8人以上の世帯等

本調査のサブサンプル中のレコードのうち、世帯人員が8人以上の世帯に係るものは、匿名データから削除する計画である。

また、本調査は、平成14年以降については、15歳未満の世帯員の人数が、0～3歳、4～6歳、7～9歳、10～12歳、13～14歳の階級別に把握されているため、同一年齢階級に3人以上いる世帯のレコードは、匿名データから削除する計画である。

これらについては、世帯員の人数は世帯の外部から比較的容易に把握可能な属性であり、出現頻度が低い世帯構成の場合、調査客体が特定される可能性が生じることから、適当である。

(イ) 自衛官、受刑者、並びに死亡・転出のレコード削除

本調査のサブサンプル中のレコードのうち、自衛官、受刑者のレコード、並びに死亡・転出した世帯員のレコードは、匿名データから削除する計画である。

これについては、自衛官、受刑者という特異な属性のレコードであること、死亡・転出した世帯員のレコードは当該月の就業状況等の内容が含まれていないものであることから、適当である。

(2) 識別情報の階級区分の統合

ア トップコーディング

(ア) 高齢者の年齢

世帯員の年齢については、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。）を行うこととし、当該上限値は85歳以上とする計画である。

これについては、出現頻度が低い一定年齢以上の高齢者をトップコーディングすることにより、他の属性情報との組み合わせによる調査客体の特定を防ぐことから、適当である。

(イ) 月末1週間に仕事をした時間

「月末1週間（ただし12月は20～26日）に仕事をした時間」については、90時間以上のデータをトップコーディングする計画である。

これについては、出現頻度が低い長時間労働者をトップコーディングすることにより、他の属性情報との組み合わせによる調査客体の特定を防ぐことから、適当である。

ただし、当該トップコーディングを行った変数については、その平均値等を匿名データの提供に併せて提供することにより、利用者の利便性向上を図る必要がある。

イ リコーディング（分類区分の再付与）

(ア) 15歳以上の世帯員の年齢

15歳以上の世帯員の年齢（トップコーディングを行う高齢者を除く。）については、その分類の程度を粗いものにする措置（以下「リコーディング」という。）を講じることとし、5歳階級別とする計画である。

これについては、各歳別のデータ提供に比べて匿名データの有用性が低下するものの、各歳別の年齢が明らかになると、「事業の種類（産業）」等の世帯員に関する他の属性情報との組み合わせにより調査客体が特定される可能性が生じることから、やむを得ない措置である。

(イ) 事業の種類（産業）**←及び**本人の仕事の種類（職業）

事業の種類（産業）**←及び**本人の仕事の種類（職業）については、公表統計の表章区分に合わせてリコーディングを行う計画である。

これについては、公表統計の表章区分との整合性を図ることにより匿名データの利便性が確保されること、詳細な産業区分、職業区分が明らかになると他の属性情報との組み合わせにより調査客体が特定される可能性が生じることから、適当である。

(ウ) 15歳未満の男女別総数

15歳未満の男女別総数については、男女の区別をしないで総数に置き換える計画である。

これについては、15歳未満の男女別の構成は外部から比較的容易に把握可能な属性であり、出現頻度の低い構成の場合に、他の属性情報と組み合わせることによって調査客体が特定される可能性が生じることから、適当である。

(3) その他の匿名化措置

ア 匿名データの提供時期

本調査の匿名データの提供時期については、調査実施後3年以上経過したものを提供する計画である。

これについては、本調査では、同一の住戸に居住する世帯に対して2年にわたり同一の2か月を調査するという調査方法が採用されていることを踏まえた上で、時間の経過により調査データと外部情報との照合を困難とする措置であることから、適当である。

3 今後の課題

(1) 特定調査票に係る匿名データの作成

本調査は基礎調査票及び特定調査票の2種類で構成されているが、本計画による匿名データの作成は基礎調査票のみが対象となっている。

特定調査票は、我が国の就業・不就業の状況を分析するための非常に有益な情報を有するものであることから、その匿名データの作成についても検討する必要がある。

(2) 匿名データの作成対象年次の拡大

本計画では、匿名データの作成対象調査を平成元年以降のものとしている。

しかしながら、研究には経年的な分析が重要であり、我が国の経済状況が大きく変化した1980年代の分析の重要性を鑑みれば、平成より前のデータの匿名データの作成及び提供についても検討する必要がある。

(3) 複数の匿名データの作成の可能性の検討

本計画では、匿名性を確保するため、地域区分や世帯員の年齢等の調査客体の特定につながる可能性がある重要かつ基本的な属性情報については、厳格な匿名化措置を講じることとしている。

しかしながら、匿名データの利用者のニーズについては様々なものが考えられ、例えば、15歳以上の世帯員の年齢を各歳別とする一方、産業、職業等の分類区分を大括り化した匿名データの作成についてのニーズも指摘されている。

調査客体の匿名性は、複数の匿名化措置により全体として確保されるものであるため、匿名化措置の内容や組合せを変えることにより、同一の調査について複数の匿名データを作成できる可能性はあると考えられることから、今後、複数の匿名データのマッチングによる調査客体の特定の危険性に関する研究等の結果や匿名データの利用者のニーズを踏まえて、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必要がある。