

「労働力調査に係る匿名データの作成について」の論点

平成 23 年 7 月 4 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

労働力調査（平成元年 1 月～平成 19 年 12 月）（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。

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 リサンプリングの方法等

リサンプリングは、地域 11 ブロック及び組符号 8 区分による層化を行った後、世帯を単位としてまとめた上で等確率抽出により、世帯を単位とするリサンプリング率が約 80%となるようにするとともに、沖縄県については、リサンプリング率を約 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としているが、匿名性の観点からリサンプリングの方法及び抽出率は適当か。

また、公表統計との間で、代表的な項目の平均値や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。リサンプリングされたレコードに対して匿名データ用の乗率を再計算・再付与する必要はないか。

イ 地域区分、事業の種類（産業）及び本人の仕事の種類（職業）

本調査は、地域区分については全国 1 区分とし、産業及び職業については報告書の表章区分に合わせてリコーディングする計画であるが、匿名性及び有用性の観点から適当か。労働市場の需給度合が地域間で異なることを鑑みると、地域間の違いを考慮した分析のニーズは高いと思われるが、例えば、産業、職業等を大きくくり化すること等により地域に関する情報を提供することは考えられないか。

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 同一年齢の子供の数

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を削除することとしているが、匿名性及び有用性の観点から適当か。例えば、年齢階級をリコーディングすること等により削除対象の世帯数が減る可能性があるが、年齢階級を見直す必要はないか。

オ 15 歳以上の世帯員の年齢

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、85歳以上の世帯員について、トップコーディングを行うことは適当か。

カ 月末1週間に仕事をした時間

月末1週間に仕事をした時間について、90時間以上でトップコーディングをする計画であるが、外観識別可能性及びデータ分析の有用性の観点からトップコーディングを行うことは適当か。

キ 前月欄の情報

本調査は、同一世帯を2ヶ月連続して調査することから、その集計上、2ヶ月目のデータには前月の「月末1週間に仕事をしたかどうかの別」、「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」の情報が保持されている。この前月欄の情報については、個人の特定化につながる恐れがあることから、「月末1週間に仕事をしたかどうかの別」のみ提供することとしているが、匿名性及び有用性の観点から適当か。例えば、前月の「事業の種類(産業)」等の情報を大きくくり化すること等により、これらの情報を併せて提供することは考えられないか。

ク その他の匿名化措置

(ア) 15歳未満の男女別総数は、男女の区別をしないで総数に置き換えることとしているが、匿名性及び有用性の観点から適当か。

(イ) 自衛官、受刑者、並びに死亡・転出のレコードは削除することとしているが、適当か。

(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 15歳未満の世帯員

15歳未満の世帯員については、平成14年以降、2～4歳階級別に人数が調査されている。これについては、親の就業状況に影響を与える特に大きな要因と考えられることから、更にリコーディングする等の匿名化措置を講じる計画とはなっていないが、匿名性の確保の観点から適当か。

イ 他の外観識別可能な事項

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性を確保すべき事項は存在しないか。

2 他の情報との関係

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) 匿名データの提供時期及び作成対象年

今回、匿名データの提供時期を調査実施後 3 年以上経過したものから提供することとしているが、匿名性及び有用性の観点から適当か。また、平成元年以前のデータについても提供することはできないか。

(2) トップコーディング等が行われた変数

トップコーディング等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等を提供することはできないか。

(3) 特定調査票の匿名データ化

労働力調査は、基礎調査票及び特定調査票の 2 種類で行われている。今回、匿名データの作成が計画されているのは基礎調査票のみであり、特定調査票については対象となっていないが、有用性の観点から匿名データの作成はできないか。

また、将来の特定調査票の匿名データ化を想定した場合の、本計画の基礎調査票の匿名データとの関係をどのように考えるか。

(4) 同一世帯のマッチング

労働力調査は 2 年にわたり同一の 2 か月を調査するものであり、理論的には前月又は前年のデータを用いて同一世帯のマッチングが可能と思われるが、有用性の観点からパネル化データを作成することはできないか。

(5) 匿名化技法の検討

中長期的には、提供可能な項目の増加等提供の幅を広げる観点から、ノイズの付与やスワッピング等の匿名化技法について、検討を行うことが必要ではないか。

「労働力調査に係る匿名データの作成について」の論点 ~~(案)~~

平成 23 年 ~~7~~月 ~~17~~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

労働力調査（平成元年 1 月～平成 19 年 12 月）（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。

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 リサンプリングの方法等

リサンプリングは、地域 11 ブロック及び組符号 8 区分による層化を行った後、世帯を単位としてまとめた上で等確率抽出により、世帯を単位とするリサンプリング率が約 80%となるようにするとともに、沖縄県については、リサンプリング率を約 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としているが、匿名性の観点からリサンプリングの方法及び抽出率は適当か。

また、公表統計との間で、代表的な項目の平均値や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。リサンプリングされたレコードに対して匿名データ用の乗率を再計算・再付与する必要はないか。

イ 地域区分、事業の種類（産業）及び本人の仕事の種類（職業）

本調査は、地域区分についてはを全国 1 区分とし、産業及び職業については報告書の表章区分に合わせてリコーディングする計画であるが、匿名性及び有用性の観点から適当か。何らかの地域表章は可能か。労働市場の需給度合が地域間で異なることを鑑みると、地域間の違いを考慮した分析のニーズは高いと思われるが、例えば、産業、職業等を大きくり化すること等により地域に関する情報を提供することは考えられないか。

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 同一年齢の子供の数

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を削除することとしているが、匿名性及び有用性の観点から適当か。例えば、年齢階級をリコーディングすること等により削除対象の世帯数が減る可能性があるが、年齢階級を見直す必要はないか。

オ 15 歳以上の世帯員の年齢

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

(見え消し版)

た、85歳以上の世帯員について、トップコーディングを行うことは適当か。

カ 月末1週間に仕事をした時間

月末1週間に仕事をした時間について、90時間以上でトップコーディングをすることは適当か。計画であるが、外観識別可能性及びデータ分析の有用性の観点からトップコーディングを行うことは適当か。

~~キ 事業の種類(産業)及び本人の仕事の種類(職業)~~

~~産業及び職業については、報告書の表章区分に合わせてリコーディングすることとしているが、匿名性及び有用性の観点から適当か。~~

キタ 前月欄の情報

本調査は、同一世帯を2ヶ月連続して調査することから、その集計上、2ヶ月目のデータには前月の「月末1週間に仕事をしたかどうかの別」、「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」の情報が保持されている。この前月欄の情報については、個人の特定化につながる恐れがあることから、「月末1週間に仕事をしたかどうかの別」のみ提供することとしているが、匿名性及び有用性の観点から適当か。例えば、前月の「事業の種類(産業)」等の情報を大きくくり化すること等により、これらの情報を併せて提供することは考えられないか。

クケ その他の匿名化措置

- (ア) 15歳未満の男女別総数は、男女の区別をしないで総数に置き換えることとしているが、匿名性及び有用性の観点から適当か。
- (イ) 自衛官、受刑者、並びに死亡・転出のレコードは削除することとしているが、適当か。

(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 15歳未満の世帯員

15歳未満の世帯員については、平成14年以降、2～4歳階級別に人数が調査されている。これについては、親の就業状況に影響を与える特に大きな要因と考えられることから、更にリコーディングする等の匿名化措置を講じる計画とはなっていないが、匿名性の確保の観点から適当か。

イ 他の外観識別可能な事項

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性を確保すべき事項は存在しないか。

(見え消し版)

2 他の情報との関係

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) 匿名データの提供時期及び作成対象年

今回、匿名データの提供時期を調査実施後 3 年以上経過したものから提供することとしているが、匿名性及び有用性の観点から適当か。また、平成元年以前のデータについても提供することはできないか。

(2) トップコーディング等が行われた変数

トップコーディング等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等を提供することはできないか。

(3) 特定調査票の匿名データ化

労働力調査は、基礎調査票及び特定調査票の 2 種類で行われている。今回、匿名データの作成が計画されているのは基礎調査票のみであり、特定調査票については対象となっていないが、有用性の観点から匿名データの作成はできないか。

また、将来の特定調査票の匿名データ化を想定した場合の、本計画の基礎調査票の匿名データとの関係をどのように考えるか。

(4) 同一世帯のマッチング

労働力調査は 2 年にわたり同一の 2 か月を調査するものであり、理論的には前月又は前年のデータを用いて同一世帯のマッチングが可能と思われるが、有用性の観点からパネル化データを作成することはできないか。

(5) 匿名化技法の検討

中長期的には、提供可能な項目の増加等提供の幅を広げる観点から、ノイズの付与やスワッピング等の匿名化技法について、検討を行うことが必要ではないか。