

統計委第 2 号  
平成 28 年 4 月 26 日

総務大臣  
山本 早苗 殿

統計委員会委員長  
西村 清彦

諮問第 87 号の答申  
就業構造基本調査に係る匿名データの作成について

本委員会は、総務省が作成を予定している平成 19 年就業構造基本調査（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、本調査の回答者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。

ただし、以下の「2 理由等」で指摘した事項については、修正が必要である。

2 理由等

(1) 世帯員の年齢の匿名化措置

本計画では、平成 14 年就業構造基本調査の匿名データ作成方法同様、年齢について 5 歳階級でまとめた上で、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。）をとり、上限値を 85 歳とすることとしている。

これについて、年齢の上限値を引き上げるべきか検討した結果、高齢者層における有業者数は少ないことから匿名性を確保するためには 85 歳以上でトップコーディングすることは適当であると考える。

(2) 本調査において変更された項目の匿名化措置

本計画では、本調査における「職業訓練・自己啓発の有無」や「起業の有無」などの新規調査項目について、調査票の回答内容を加工せずにそのまま提供することとしている。これについては、有用性が高まることから、適当である。

ただし、「元号、年、月」の情報で提供される「初職に就いた時期」が相当程度経過している者については、サンプルサイズが小さいことに加えて、特に初職

に就いて70年以上経過した者については、その「初職に就いた時期」から調査時の年齢の下限値が推測されることにより、85歳以上をトップコーディングした効果を損なうおそれがある。そのため、初職に就いた時期に関連する項目のうち、現職が初職の場合は「就業開始時期」、現職・前職以外が初職の場合は「初職に就いた時期」をそれぞれトップコーディングし、前職が初職の場合は「前職の継続就業期間」を必要に応じて置き換えた上で提供すべきことを指摘する。

また、「初職に就いた年齢」は提供しないこととしているが、提供した場合は特定の年齢が明らかとなる場合があるため、匿名性を確保するために適当であると考えられる。

(3) 前回答申（諮問第13号、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

ア 複数の匿名データ作成の可能性

前回答申において、就業構造基本調査の匿名データに関して、年齢を各歳別にしたファイル作成が「今後の課題」として挙げられている。これについて総務省は、リサンプリング率8割で世帯を単位として抽出した第1のファイルと、そこから年齢を各歳別にした個人を単位とした小さな抽出サイズの第2のファイルを作成し、両者を属性項目でマッチングすることで匿名性の確保を検証した。その結果、マッチングに利用する項目を増やすことで匿名性の確保ができなくなる危険性があることが判明したため、この方法による第2のファイル作成は見送り、別の方法による作成可能性について引き続き検討することとしている。

これについて、検証した方法によるデータ作成では匿名性が十分に確保できない可能性があることから、現段階においては複数ファイルの作成を見送ることはやむを得ないものとする。ただし、上記の「別の方法」については、後述「3 今後の課題」で示した方向で検討する必要がある。

イ 匿名データ提供時期の短縮化

前回答申において、調査実施後5年以上経過したものという提供基準について、年数の緩和を検討することが「今後の課題」として挙げられている。これについて総務省は、当該調査が5年周期であり、次回調査を実施する際に、直近の調査結果が匿名データとして提供されることで、調査対象者の不安感が募る懸念があることから、調査実施後5年以上経過してから提供することとしている。

これについては、調査環境等への影響を考慮すれば、やむを得ないものとするものの、平成元年より前の年次拡張や直近年次の早期提供に向け、引き続き努力すべきと考える。

ウ トップコーディングが行われた変数の平均値等の整備

前回答申において、トップコーディングした変数について、多変量解析等

に資するため、一定程度の情報を提供することが「今後の課題」として挙げられている。これについて総務省は、85歳でトップコーディングした層の平均年齢等を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ので、適当である。

### 3 今後の課題

就業構造基本調査の匿名データは、平成4年、9年、14年に続き本調査で4年次分のファイルが提供されることになる。同調査は、女性の結婚、出産等に関連した就業動向や、若年、中高年齢者の退職・再就職等の動向など近年の雇用動向の変化を分析する上で貴重な情報源であり、このような観点から分析する上では、年齢や地域等に関する詳細な情報が不可欠である。同調査については、この間、利用者の幅が広がってきており、年齢区分、産業・職業の分類区分、地域区分の細分化など利用者のニーズが高度化してきているとの指摘もある。さらに、利用者の使い易さに配慮を求める意見もある。

一方、匿名性は、一つの匿名化措置のみで確保されるのではなく、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することが可能であると考えられる。そのため、分析上のニーズを的確に把握し、そのニーズに対応した匿名化措置の内容や組合せ、抽出単位の異なる複数ファイルの作成可能性について検討する必要がある。さらに、複数ファイルの第2のファイル作成に関しては、すでに提供されている匿名データとは独立してリサンプリングする方法などを検討するとともに、匿名性の確保に留意しながら、年齢あるいは地域の詳細な情報を含む匿名データの提供の在り方を引き続き検討する必要がある。