

第21回 匿名データ部会 議事概要

1 日 時 平成28年4月19日（火） 10:00～10:55

2 場 所 総務省第二庁舎6階 特別会議室

3 出席者

(部会長) 北村 行伸

(委員) 川崎 茂、永瀬 伸子

(専門委員) 南 和宏、村田 磨理子

(審議協力者) 総務省（政策統括官（統計基準担当））、財務省、文部科学省、経済産業省、
農林水産省、国土交通省、東京都、千葉県

(諮問者) 総務省：統計局統計調査部調査企画課 北原課長ほか

(事務局) 総務省統計委員会担当室：佐々木室長補佐ほか

4 議 事

(1) 就業構造基本調査に係る匿名データの作成について

(2) その他

5 議事概要

総務省から前回部会で検討事項とされた「初職に就いた時期」の扱いについて説明した後、匿名データ作成・提供について審議が行われた。その後、答申（案）の審議が行われ、所要の修正を行うことで採択された。なお、答申（案）の修正については部会長に一任することとされ、平成28年4月26日に開催予定の第97回統計委員会において、部会長から報告することとされた。

委員等の主な意見は以下のとおり。

(1) 初職に就いた時期の扱い

総務省の説明

- ・「初職に就いた時期」は、前回の部会では報告書への掲載内容と混同して説明したが、匿名データでは「元号、年、月」で提供することとしている。データの時期別分布を精査した結果、初職に就いた月は、4月に6割程度が集中しているものの1月から12月にもばらついており、年月で提供しても匿名化は確保できるものとする。ただし、初職に就いた時期が昭和12年以前の者も少数おり、この場合、調査時点から70年以上であり、初職に就いた時期から調査時の年齢の下限を推測された場合、85歳でトップコーディングした匿名化措置の効果を損なうおそれがある。そのため、初職に就いた時期が70年以上前の場合は、70年となるようにトップコーディングすることを考えている。具体的には、調査項目との関係で、初職が現職、初職が前職、初職がそれ以外の3者で異なる項目に適用される。

質疑等

- ・70年以上前のことを年月単位で正確に回答できているのかということも考慮すると、70年以上をトップコーディングすることは適当である。
 - ・70年以上をトップコーディングするのは適切な処置である。前職が初職の場合は、初職に就いた時期が不明の場合もあることを注意喚起する必要がある。
 - ・年月で提供されることは有益である。平成14年の匿名データの提供方法と一部異なる部分があるということか。
- そのとおり。平成19年調査での設問項目「最初について仕事は現在の仕事又は前の仕事と別ですか（C8）」が追加されたため、現職や前職が初職かどうか半別できるようになったことの影響である。
- ・提供される情報が従来と異なることを理解できるようにしてほしい。
 - ・初職に就いた時期が年月まで提供されることはよいことだ。ただ、その年月は、初職が現職とその他の場合は直接的に求められ、前職の場合は計算して求めることになり、後者は合成変数であることを利用者に提示する必要がある。4月に初職に就いた割合が昭和40年代から60年代にかけては高かったが平成10年以降年々低下していることが分かるなど貴重な情報だと思うが、地域情報があると労働需給との関係が分かりもつと有益である。
 - ・提供している情報が、調査票の生データなのか加工処理したデータなのかを示すことが必要。加工処理して提供する場合は、そのメリットを吟味する必要がある。

部会長のまとめ

- ・初職に就いた時期について、「元号、年、月」で提供することは適当であるが、年齢との関係から必要に応じて一部をトップコーディングすべきである。

(2) 答申（案）

- ・複数の匿名データ作成の可能性のところで、8割のリサンプリングの中から、次の各歳別のデータを抽出しようとするマッキングのリスクが高まるので、もう1回別に8割をリサンプリングした上で、その中から抽出したほうがよいという議論があった。それを踏まえると、「リサンプリング率8割の妥当性を含め」というよりも、改めてリサンプリングしたものの中から各歳別のデータを提供することを含めてというイメージで表現していただきたい。
- ・同じ箇所でも、前回の部会での説明は分かり易かった。8割の中からはとるか、重なる形ではとるか、また、別に抽出するか、3つのケースがあって、そのうち8割の中からはとったケースは危険の可能性が検証できたが、その他の2つのケースに関しては、安全性の検証を確立するところまではできなかった。それを踏まえたような文章にしてほしい。
- ・サンプリング手法を使った匿名化措置によって、安全性が大幅に高められるという研究事例もかなり出てきている。サンプリング手法については、もっと研究を深めるべきかと思うが、この部会の限られた期間では議論が難しいので、何らかの継続した議論の場があれば、より意味のある研究ができるのではないかと思う。

- ・就業問題を深く分析するためには、地域という変数が非常に重要であり、そういうニーズに対応するためにも2つ目のファイル提供が必要であるというような表現ができないか。
- ・「初職に就いた時期」から年齢の下限値が推測されることの表現が分かりにくい。
- ・匿名データについて、各歳別情報がないことと地域情報がないことは使い勝手を悪くしていると思う。このため匿名データの利用が33条申請に比べて伸びていないと思われる。データが何十万件あるというよりは、もう少し小さくてよいので、地域別のバリエーションがある方がよい。女性の就業構造は、年齢各歳別、出産時期により変わってくるし、若年の就職時期についても、学校を卒業してから1年間無職期間があったのか、半年で就いたのか、1か月後に就いたのかというのは労働市場を分析するには重要な情報であるが、今回の匿名データでは、卒業後どのくらい無職期間があったのか分からない。そのため、地域、年齢が分かるファイルを提供してほしい。
- ・海外からもデータを利用し易くなるように検討してほしい。
- ・匿名データ全般に係ることだが、地域区分の提供は有益とは思いますが、ミニマムどれくらいの粗さの地域区分が必要なのか、難しい。細かく言えば、市町村レベルかもしれないが、県単位でよいのか3大都市圏、4大都市圏、それ以外でもよいのか、ユーザからも意見を出すようなことがあってもよいのではないかなと思う。そういう議論ができればよい。
- ・今後の課題の書きぶりで、利用者のニーズが多様化してきていることをもう少し具体的に記述したほうがよい。前回の答申は、年齢を各歳別にする一方で、職業、産業別の分類区分を大きくりにするという内容で、今回は、逆に産業、職業分類を細かくするとか、地域区分を出すとか、色々なバリエーションがあるということを示しているが少し具体性に欠ける。

部会長のまとめ

- ・答申案については、いただいたご意見を踏まえて修正する。
- ・匿名データを最初に作成したときは、全国消費実態調査など4つの調査をまとめて検討されたので、地域や年齢のトップコーディングなどが共通的に措置されている。その後の状況としては、調査に応じた提供の仕方を考えないといけない時期になっていることを感じている。特に、地域差の特徴を利用して政策の効果の違いをみることができないので、地域情報の提供方法は改善を期待したい。
- ・上記のような匿名化手法の検討については、匿名データ部会で継続的に検討するか共通の課題として統計委員会に対応することを考えたい。

以上

<文責 総務省統計委員会担当室 速報のため事後修正の可能性あり>