

第9回匿名データ部会 議事概要

1 日 時 平成23年8月1日(月) 14:57~15:51

2 場 所 中央合同庁舎4号館2階 共用第3特別会議室

3 出席者

椿広計部会長、井伊雅子部会長代理、廣松毅委員、伊藤伸介専門委員、稲葉由之専門委員、安田聖専門委員、総務省(政策統括官(統計基準担当))、厚生労働省、農林水産省、経済産業省、国土交通省、日本銀行、東京都、千葉県

【諮問者(総務省統計局統計調査部)】

高田統計審査官、北林調査企画課長補佐、横内調査企画課二次利用推進係長、佐藤国勢統計課労働力人口統計室企画官

【事務局(内閣府統計委員会担当室)】

若林参事官、高部参事官補佐

4 議事次第 (1) 労働力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

(1) 答申(案)の審議について

椿部会長から、資料「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)』」が示され、事務局が答申(案)について項目ごとに朗読の後、審議が行われた結果、一部で所要の修正を行うこととされたが、答申(案)は概ね適当であるとして採択された。

なお、答申(案)の修正分の表現については部会長に一任することとされた。

各委員等の主な意見等は、次のとおり。

ア 識別情報の削除等

- ・ 労働力調査は、同一世帯を4回調査するというパネル的な構造を持っているため、世帯のパネル構造に係る情報が与えられずに分析すると、誤差項の系列相関があるにもかかわらず、系列相関を無視した分析になってしまう。このため、むしろ一度抽出された世帯を二度と抽出しないようリサンプリング方法とし、対応関係のないデータの形での匿名化は考えられないかという指摘があった。

指摘された方法によりリサンプリングを行うと出現率を4分の1にしなければならず、約20%のリサンプリング率となる。現行の8割リサンプリングであれば、オリジナルデータと

平均値や分散などの統計量はかなり一致しているが、指摘された方法によりリサンプリングを行うと、サンプル自体が非常に少なくなってしまい、分析上いろいろな統計量がオリジナルのデータとは若干ずれてくる可能性がある。

- ・ この問題は、リサンプリングの話ではなく労働力調査そのものの特性であり、逆に同一世帯を二度抽出しないようにすると、労働力調査とは異なるデータセットになるのではないか。匿名データの作成方法と労働力調査の特質を分けて考えた方がよい。

部会長のまとめ

- ・ 基本的に、対応のあるデータの中から一つだけを選んで、データの独立性の方を担保すべきという意見と考えるが、それは匿名データの規模を4分の1にすることとなり不安定性が増すのではないか。また、本来のデータとは異なるものになる可能性もある。仮に対応のあるデータが隠れていたとしても、一定以上の時期を離れて考えれば相関は無視できるので、むしろ、元のデータ構造を維持している原案のとおりとしたい。

- ・ 「異動」という用語が、「(ア)地理的情報の削除等」では異時点間の問題として使われており、また、「(イ)前月欄の情報削除」では「異動情報」あるいは「異動に関する情報」という使われ方をしているが、この用語でよいのか。

一般的に従業上の地位等が変わる場合に「異動」という言葉を使用するため、答申案でもこの言葉を使用した。

労働力調査の調査票では、転出・死亡で調査対象から外れる場合に「異動」という用語を使用しており、「異時点間の」とした方が紛れがない。

部会長のまとめ

- ・ 「異動」という用語の使い方について検討することとし、最終的な修文については、部会長に一任とさせていただきたい。

イ リコーディング(分類区分の再付与)

- ・ 「(イ)事業の種類(産業)及び本人の仕事の種類(職業)」の文章中に「事業の種類(産業)及び本人の仕事の種類(職業)については」とあるが、「事業の種類(産業)」と「本人の仕事の種類(職業)」はそれぞれ調査項目であり、他の箇所での書き方に合わせて、括弧(「」)を付ける必要がある。

部会長のまとめ

- ・ 括弧を付ける形で修文することとしたい。

ウ 今後の課題

- ・ 「(3)複数の匿名データ作成の可能性の検討」について、リサンプリング率が80%と高い場合には、複数の匿名データを作成することはかなり難しいが、もう一種類匿名データを作成するという事を考えているのか。

例えば年齢階層を細かくし、産業分類や職業分類をほぼ提示しない匿名データのニーズもあるため、別の匿名データを作成する可能性を示唆したもの。ただし、2種類の匿名データが提供されることにより新たなリスクも生じるため、その部分に関する課題を後段に記述している。

ご指摘のような2種類の匿名データが提供されることによる新たなリスクも含めて検討してまいりたい。労働力調査の匿名データについてどこまで細かい区分とするかは、労働力調査自体がどこまで表章しているかも含めて検討する必要があるが、以前、就業構造基本調査の匿名データの作成の際にも同様の議論があったこと、また、前回の部会においてオーダーメイド集計でもっと細かい区分が提供できないかという議論もあったことから、そのようなことも含めて、今後、総合的に考えていきたい。

- ・ 「程度が異なる」とあるが、これは、例えば分析には使用できないが教育用には利用できるように匿名データを作成する可能性について記載したものか。

ここでの「程度」とは、個別の匿名化措置の程度を指している。

- ・ 今回の匿名データで研究者のニーズに対応できない部分については、統計法第33条の調査票情報の提供で対応されるのか。

科研費を取得する等、高度な公益性のある研究等の要件を満たしていただければ、33条による調査票情報の提供により積極的に対応したいと考えている。

部会長のまとめ

- ・ 「今後の課題」として3つ挙げているが、原案のとおり了承されたこととしたい。

エ その他の意見

- ・ 「調査客体」という用語が何カ所かで使われているが、リサンプリングは世帯単位で行われているため、ここでいう調査客体は「世帯」と解釈するということが。

答申(案)では、個々の世帯員が特定化されるリスクという意味でも「調査客体」が使用されており、世帯並びに世帯員であると解釈するのではないか。

部会長のまとめ

- ・ 労働力調査のような世帯調査の場合、調査の対象となったのは世帯並びにその構成員である世帯員であるため、匿名性の確保の観点から、「調査客体」については世帯並びに世帯に属する世帯員を含むという解釈をすることとしたい。

(2) その他

- ・ 本日は承された答申(案)及び議事概要については、8月29日(月)の統計委員会に提出することとされた。

以上

<文責 内閣府大臣官房統計委員会担当室 速報のため事後修正の可能性あり>