

2016年1月8日
美添泰人

国民生活基礎調査の検討課題

比推定に関する厚労省の考え方は、特殊と思われる。比推定とは、一般に補助変数 Y を利用して目的とする変数 X の推定効率を高める手法と理解する必要がある。(一般的な記号は、厚労省の記述とは X, Y が逆の場合もある。)

現行の方法では、都道府県別推計人口を利用した比推定が世帯票関連の変数(ある属性の世帯数など)を推定するために用いられている。世帯類型別(単身・二人以上)に回答率が異なるという非標本誤差に対応する手法ではないため、何らかの対策が必要である。

12月28日の部会では、「世帯類型・都道府県別(この理解は正しいか?)」に層別した推定結果(岩崎研究会の報告書所収)が紹介された。その際の説明では「世帯数が一致するのは合わせに行っているのだから当然」であり、一方「年齢階級別の推定結果は現行の手法より大幅に悪化する」ということであった。

この点も含めて、以下の点を確認したい。

1. 比推定の効果を評価するために対象とする変数については、より広い範囲で検討すべきである。「世帯数が一致するのは当然」と言いながら、完全に一致しているわけではない。精度が良い理由は、総世帯数と人口の間に比例に近い強い関係があることで、これが比推定の理論的根拠である。
関心の対象である(いくつかの属性の)世帯数 X についても、人口 Y との間に近似的に比例関係が存在することは確認されていると考えるが、そのような資料があれば、提示して欲しい。比推定の効率は目的とする変数によって異なる。
2. 「世帯類型・都道府県・年齢階級」別に層化して推定したのだろうか。そうであれば結果を知りたい。これらの層で、国勢調査の世帯数を補助変数として比推定を用いれば、少なくとも回収率の差による非標本誤差は、相当程度減少されると予想される。
3. 「年齢階級別で誤差が大きいことが反例だから、他の検討は不要」という理解は正しくない。統計学では確率的な挙動が問題であり、推定の精度を評価しない限り、仮にある推定結果が良くないとしても、証明にはなっていない。必要な情報は各推定方法に関する誤差の評価および適当なモデルの下での非標本誤差の評価である。
4. 現行の比推定によっては、非標本誤差の解決は得られない。他の対策を検討する必要があることは明らか。
5. 社会福祉施設等に住む約100万人の処理次第で、現行の手法がより優れている根拠のひとつとした人口の妥当性はまだ検討の余地がある。

他の点に関する個人的意見は以下のとおり。

- 集落抽出は悪い方法ではないと考えているが、それでも、単純無作為と比較して、どれほど効率が低下するかは評価する必要がある。
- 前回の議論で、国勢調査の調査区を厚労省があらかじめ抽出して世帯票の調査対象とする手順について、調査区によってマンションの新設などで世帯数に大きな違いが発生すると調査が困難になるとの説明があった。その困難を解決するためには、各調査区から一定の世帯を抽出すれば良いと指摘されたら、どのように回答するかは準備しておいた方が良いと思う。
- 国の調査を民間に委託する場合の例として「国勢調査区を抽出枠とする無作為抽出」という仕様書は少なくない。そのような調査に特段の困難はないと指摘された場合、集落抽出を正当化するための回答も必要だろう。

- 岩崎研究会のような検討は、継続的に実施することが望ましい。実際、当時の研究会委員に、現時点で同じ問題の検討を依頼したら、さらなる改善点、検討課題が提示されるものと思う。
- 他の府省と協力して、これらの課題を検討する機会を設けてはどうだろうか。