

国立社会保障・人口問題研究所 ディスカッションペーパー
「2022 年生活と支え合いに関する調査世帯票のウェイトについての考え方と方法」
に関するコメント

2024 年 3 月 29 日
国立社会保障・人口問題研究所
榊原 賢二郎

(1)事後層化ウェイト

度数 0 でウェイトが計算できなかった層があるようですが、レイキング(反復比例推定)したらいかがでしょうか。生活と支え合いに関する調査で得られたクロス表の行・列全体がゼロ度数でない限りは、度数ゼロのセルがあってもウェイトは計算できることになります。

(2)信頼区間の計算方法

ウェイトをかけた信頼区間の計算について、平均から標準偏差の 1.96 倍という方法が採用されていますが、ブートストラップ法で信頼区間を計算しても同様でしょうか。

(3)平均世帯人員数

既に国民生活基礎調査の人員数を無回答ウェイトの中で統制しており、ウェイト 2 でも世帯数を使っています。もし可能であれば他の評価指標があった方が良いのですが、項目無回答の件で厳しいのでしょうか。

(4)層化

今回は都道府県による層化(あるいは多段抽出の一段目)と認識しております。世帯統計室では、人口集中・非集中などで更に層化していた気がしますが、そこは抽出率が同じであるという仮定を置いているということになりますでしょうか。

この手の特性を活用しているあたりで、多段抽出というよりは層化集落抽出という表現になりますでしょうか。論理上、都道府県(+人口集中・非集中)を層とする層化集落抽出は、1 段目と 3 段目の抽出率が 100%の多段抽出ということで、同じになるかもしれませんが。

(5)個人票への適用

個人票に関するウェイトについて、ご教示いただければと存じます。

・ベースウェイトについては、少なくともベースウェイト 1 は使えるという認識でよろしいでしょうか。

・無回答ウェイトについてですが、世帯票のロジットと同じ処理(世帯人員数不明ダミーの導入)が通用しないということでしょうか。(e.g. ポスティング世帯を世帯員数不明世帯に追加で投入するなど)

(6)都道府県集計と不等確率抽出

今回検討していただいたウェイト算出方法を使えば、不等確率抽出が可能になります。現在でも、端数の問題で、小幅に不等確率となりますが、さらに大幅な乖離が可能となると考えられます。この点を、都道府県集計に活用できないかと考えております。

生活と支え合いに関する調査のような、公的統計の中では必ずしも大規模ではない調査について、全国集計だけでなく都道府県集計も可能にするためには、調査地区数の拡大が必要となります。ただ、等確率抽出に固執すると、膨大な調査地区数の追加が必要となり、現実的ではありません。仮に、誤差率を一定以下に抑えるために、各都道府県に 15 調査地区以上必要だとすると(この地区数の計算は別途の課題として認識しております)、等確率抽出の場合、これまで(1 調査地区のみの県が存在するとします)の 15 倍の地区数となります。これに対して、各都道府県一律 15 調査地区とすれば、今回の 2 倍強の地区数となります。このような不等確率抽出が承認されうるかどうか、その他いかなる障壁がありうるかについて検討しておく必要があると思われまます。

私の方で一点気にかかるのが、回収率の計算です。基本的に大都市部の比率を下げることでありますが、この場合、単純に回収率を計算すると、等確率抽出に対して過大評価になるという問題があります。それでは回収率にもウェイトをかけるのか、ウェイトをかける場合、ベースウェイトのみかどうかという点について、何か知見があればご教示いただければ幸いです。