

# 社人研ディスカッションペーパー

## 「二段階調査データの復元推計」に対するコメント

2025年11月26日

横浜市立大学データサイエンス学部長 土屋隆裕

統計調査の集計においては、得られた集計値のみならず、それを導いた過程を整備し説明可能にすることが重要である。調査ウェイトについても、結果的に単純集計と復元推計がほぼ変わらない場合もあるが、適切な作成方法を採用することに意義がある。本ディスカッションペーパーには、ウェイトを丁寧に作成しようとする意図が見受けられるが、より良いウェイトの作成のために、以下の通りコメントする。

### 1. 無回答ウェイト

今回の原稿では、キャリブレーション法により世帯無回答ウェイトの計算を省略しているが、キャリブレーションと無回答ウェイトでは活用できる変数が異なるため、二相法を採用するのが良いのではないか。調査員の実査由来情報以外に、国勢調査の小地域集計を用いて、調査地区の特性を利用することができる。例えば、地区の大卒者比率(無回答者は小中高卒と考え、分母に無回答者を含む)は、補正に有用であるとされている。他に産業構成なども調整に使用することができる。

### 2. 多重代入

多重代入を行っているのは良いと思う。仮に計算量等の問題で単一代入に戻る場合、回帰代入以外に、予測平均マッチング等のマッチング法を用いることも検討していただくとよい。

### 3. 誤差評価

誤差評価の方法には、線形化とリサンプリングがあるが、線形化は無回答ウェイトがもたらす誤差を十分反映しないため、リサンプリングを採用するのが望ましいように思われる。ジャックナイフ法は、一つずつ要素を除外していけば良く、計算が比較的単純であるが、中

中央値等分位数の計算には適切ではない。そうした観点では BRR の活用が考えられる。ただし、直交行列を用いるなど、より複雑になる。多重代入と組み合わせると、計算量が更に増大するという問題があるが、検討してはどうか。

#### 4. 実査に伴う変数

居住住居形態を補助変数としてキャリブレーションするという提案があるが、実査情報の水準で、調査員が外観から無回答補正のための手掛かりを集めることもできるのではないか。オートロック有無、表札有無(プライバシー意識とも関連している)など。

#### 5. 不等加重効果

ウェイトが過大かどうかを測定する尺度として不等加重効果があり、この確認手続きを組み込んでおくのが望ましい。(deff とは異なる意味で)デザイン効果と呼ばれることもある。ウェイトの変動係数を  $cv$  とすると、 $1 + cv^2$  で表される。これを見る意義は、ウェイトの変動が大きいほど、推計値のバイアスは小さくなるものの、誤差分散が大きくなる点にある。比較対象は、ウェイトがすべて等しかった場合の誤差分散である。目安は明確に定められるものではないが、例えば 1.5 程度を一つの目安とするのはどうか。不等加重効果が過大であった場合、キャリブレーションの後の段階として、ウェイトトリミングを行うことを検討する。トリミングすると、補助変数の加重合計が母集団合計とずれてしまうので、再度キャリブレーションを行う。トリミング後に再度キャリブレーションを行うことで、不等加重効果を低減することができる。