

2019年全国家計構造調査 標準誤差の推定方法について

令和3年6月25日

総務省統計局

本議題の概要

- 2019年全国家計構造調査結果における誤差の推定方法について検討
- 前回調査までに用いてきた手法をそのまま適用することは困難
- 従来の手法に代わる新しい手法の妥当性についてご意見をいただきたい

従来の手法からの変更点

平成26年全国消費実態調査：理論式による導出



2019年全国家計構造調査：ブートストラップ法を利用した推定

想定される論点

- 復元抽出を行う区分：調査の標本設計をどの程度反映させる必要があるか
- 結果精度の評価方法：「公的統計」としての信頼度をどのようにして測るか
- 実装するにあたっての計算理論及び計算コスト
- 他の公的統計における事例

など

前回（平成26年）調査における算出方法

標本抽出の方法

各都道府県について、層化二段抽出又は層化三段抽出

- 市の場合 : 各市 → 調査単位区を抽出 → 調査世帯を抽出
- 町村の場合 : 各県内地域 → 調査町村を抽出 → 調査単位区を抽出 → 調査世帯を抽出

集計用ウエイトの作成方法

ウエイト = (各市町村における抽出率の逆数) × (比推定を利用した補正係数)

補正係数は以下の区分ごとに労働力調査結果により算出

- 二人以上の世帯 : 地方10区分 × 世帯人員階級 4 区分
- 単身世帯 : 地方 6 区分 × 男女・年齢階級 6 区分

ウエイトの合計を「調整集計世帯数」と呼ぶ（概念として母集団の世帯数とおおむね一致）



市町村間分散の不偏推定値 $\hat{\sigma}_{bbL}^2$ 、調査単位区間分散の不偏推定値 $\hat{\sigma}_{bwL}^2$ 及び調査単位区内分散の不偏推定値 $\hat{\sigma}_{wwL}^2$ をそれぞれ算出し、調整集計世帯数の比を使用して足し上げることで「地方×都市階級」別の標準誤差や、都道府県別の標準誤差を算出

調査上の変更点とブートストラップ法の採用

結果集計における前回調査からの変更点

- 4つの異なる方法で行った調査結果を統合集計
 - 基本調査 : 前回調査と同様の調査方法
 - 簡易調査 : 世帯票及び年収・貯蓄等調査票のみ使用（家計簿なし）
 - モニター調査 : 民間の調査モニターによる調査（無作為抽出ではない）
 - 特別調査 : 家計調査世帯に追加調査を実施した上で調査結果を利用
- 集計用ウエイトの作成に傾向スコアとIPF法を利用
 - モニター調査における世帯属性の偏りを補正するために、傾向スコアを利用した逆確率重み付けを適用
 - 国勢調査結果を元にした都道府県別の世帯属性分布を再現するように、IPF法を利用して集計用ウエイトを細かく補正

➡ 理論式による誤差の導出は困難

ブートストラップ法を用いたシミュレーションにより誤差を推定

※十分な数の調査データに対し、復元抽出による標本データの作成（リサンプリング）を何度も繰り返すことで、種々の推定量の分散等を求めることが可能

カナダの事例：SHS

カナダの家計調査 "SHS"(Survey of Household Spending)では
2003年以降、誤差推計にブートストラップ法を用いている

SHSとは…カナダで隔年に実施されている家計調査。

インタビューと家計簿調査を併用しており、インタビュー回答世帯に家計簿の記入をお願いしている。

	抽出方法	調査対象 世帯数	インタビュー 回答世帯数(率)	家計簿 回答世帯数(率※)
南部 10州	層化二段抽出 各層 →Geographic areas →住居	17,491世帯	10,890世帯 (62.3%)	7,566世帯 (43.3%)
北部 3準州	層化一段抽出 首都のみを対象 各準州の準州都 →住居	937世帯	590世帯 (63.0%)	364世帯 (38.8%)

表の標本設計になったのは2010年以降であり、表の数値は2019年調査のものである。
回答世帯数、回答率とは、実際に統計データを作るのに用いることのできた世帯数、率である。
※ここでの家計簿回答世帯率とは家計簿回答世帯数/調査対象世帯数である。

標本抽出方法が複雑であるため、ブートストラップ法で誤差を推計

カナダの事例：SHS

SHSでの誤差推計の手法 (Charlebois & Dubreuil, 2011)

- ① 各層ごとに (第一次抽出単位で抽出した数-1)回、第一次抽出単位の復元抽出を行う。得られたブートストラップ標本の加重平均を求める。
- ② 上記操作を1000回繰り返し、1000個のブートストラップ平均を得る
- ③ 1000個のブートストラップ平均の平均値と分散を求め、標準誤差率を求める。

この手法では、Geographic area(調査区)のみリサンプリングを行い、標準誤差率を推計している（世帯のリサンプリングは行っていない。）。

カナダの事例：SHS

SHSの具体的な手法 (Charlebois & Dubreuil, 2011)

b回目のh層のリサンプリングにおいて w_k を第一次抽出単位(PSU) i の世帯 k のウェイト、 n_h をh層で抽出したのPSUの数、 $mult_{i,b}$ をb回目のリサンプリングでPSU i が選ばれた回数とおくと、b回目のリサンプリングでの世帯 k のウェイトは次のように表される。

$$w_{k,b}^{init} = \left(\frac{n_h}{n_h - 1} \right) mult_{i,b} \cdot w_k$$

θ_k を世帯 k で報告された値とおくと、b回目の標本の平均値 $\hat{\theta}_{boot,b}$ は以下のように表される。

$$\hat{\theta}_{boot,b} = \frac{\sum_{k=1}^{n_b} w_{k,b}^{boot} \cdot \theta_k}{\sum_{k=1}^{n_b} w_{k,b}^{boot}}$$

$\hat{\theta}_{boot}$ をリサンプリングの平均値の平均とおくと、全部でB回リサンプリングを行ったときの平均値の分散 $\hat{var}_{boot}$ と標準誤差率 CV (CV は原文まま)は以下の通り求められる。

$$\hat{var}_{boot}(\hat{\theta}) = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_{boot,b} - \hat{\theta}_{boot})^2$$

$$CV(\hat{\theta}) = \frac{\sqrt{\hat{var}_{boot}(\hat{\theta})}}{\hat{\theta}}$$

平成26年調査データによる試算

標本抽出の詳細

調査市町村抽出		調査単位区抽出	調査世帯抽出
市部 791市		1市あたり 2 ～ 66 単位区 (二人以上の世帯数により比例配分)	1単位区当たり 二人以上の世帯：11 世帯 単身世帯：1 世帯
県内地域 168グループ	郡部 212町村	1町村当たり 2 単位区	

集計方法

- 簡単のため、町村抽出は考慮しない
- リサンプリング後にウエイトの比推定の再計算は行わない
- リサンプリングの方法は以下の3通り
 - ① 各市町村から単位区を復元抽出 → 各単位区から世帯を復元抽出
 - ② 大都市及び中都市のみ単位区を復元抽出 → 各単位区から世帯を復元抽出
(小都市及び町村は抽出単位区数が少ないため、形式的に悉皆層として扱う)
 - ③ 各市町村から単位区のみ復元抽出 (カナダのSHSと同様の手法)
- 2000個のブートストラップ標本を作成 (計算はR v.3.6.1 を使用)

ブートストラップ標本平均の分布からバイアス及び標準誤差率を計算

- バイアス = リサンプリング平均の平均値 - 元データの平均値
- 標準誤差率 = リサンプリング平均の標準偏差 ÷ リサンプリング平均の平均値

試算結果（平成26年全国消費実態調査結果）

消費支出（二人以上の世帯）のバイアス（円）（▲はマイナスを表す）

都市階級	集計値注	手法①	手法②	手法③
全国	292,882	86	▲ 24	▲ 18
大都市	306,579	170	▲ 60	▲ 94
中都市	291,864	87	19	1
小都市A	298,551	▲ 45	▲ 11	52
小都市B	281,404	83	▲ 26	▲ 80
町村	277,263	152	▲ 93	30

注「小都市B」及び「町村」の金額は集計用個別データから集計

消費支出（二人以上の世帯）の標準誤差率（%）

都市階級	公表値	手法①	手法②	手法③
全国	0.4	0.49	0.48	0.41
大都市	1.1	1.31	1.36	1.08
中都市	0.5	0.63	0.66	0.50
小都市A	0.6	0.79	0.51	0.67
小都市B	1.0	1.24	0.89	1.02
町村	1.3	1.41	0.95	1.12

2019年調査データへ適用する際の注意点

標本抽出の詳細

・基本調査及び簡易調査

調査市町村抽出		調査単位区抽出	調査世帯抽出
市部 793 市		1市あたり 1 ～ 41 単位区	1単位区当たり 二人以上の世帯：10 世帯 単身世帯：2 世帯
県内地域 179 グループ	郡部 215 町村	1町村あたり 1 ～ 2 単位区	

・モニター調査

標本抽出	調査世帯数の配分
約 2,800世帯を民間事業者のモニター登録者から抽出	平成27年国勢調査結果による地域（都道府県、地方）及び世帯属性（男女、年齢階級）別の単身世帯数の比率に応じて配分

・特別調査（基本調査と合わせて標本設計）

調査市町村抽出	調査単位区抽出	調査世帯抽出
市部：151 市	1市あたり 1 ～ 45 単位区	1単位区当たり 二人以上の世帯：6 世帯 単身世帯：1 世帯
郡部：17 町村	1町村あたり 1 ～ 2 単位区	

試算結果のまとめと今後の検討課題

平成26年結果の再現

- 単純に二段抽出を行うと過大推定、標本数が少ない場合を考慮した方法では過小推定
- カナダの統計調査で用いられている、調査単位区のための抽出方法で、公表値とおおむね近い値となるが、細かい区分（都道府県別、世帯属性別など）に適用可能かどうかは不明のため、理論面の整理を含めて引き続き検討が必要

推定精度向上等に向けた検討・改善案

- 町村抽出を考慮したリサンプリングの導入（三段抽出）
 - 町村抽出は二人以上の世帯数に比例した確率比例抽出を行っているため、椿・會田（2019）にあるような、無作為抽出結果との比較も検討
- 有限母集団からの非復元抽出であることを考慮したリサンプリング・集計方法
 - Rescaling Bootstrap や Mirror-Match Bootstrap の導入を検討
- リサンプリング後のウェイト補正や傾向スコアの再計算による影響の検討
 - 特に2019年では計算コストが大きく増えることに注意が必要
- モニター調査を考慮した標準誤差の推定方法の検討
- 平均値等の信頼区間の推定に関する検討
- 2019年結果の精度評価方法の検討

参考文献

- Charlebois, J. and Dubreuil, G. (2011) “Variance Estimation for the Redesigned Survey of Household Spending”, Proceedings of the Survey Methods Section, Statistical Society of Canada Annual Meeting, June 2011.
- Rao, J.N.K., Wu, C.F.J. and Yue, K. (1992) “Some Recent Work on Resampling Methods for Complex Surveys”, Survey Methodology, 18, 2, 209-217
- 統計情報研究開発センター (1998) 「標本誤差の推計方法 – 最新時代の理論と実証 –」, Sinfonica研究叢書
- 椿・會田 (2019) 「リサンプリングによる労働力調査推定精度評価」統計研究彙報 第76号p.39-50