

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4年3月1日
総務省統計局

本議題の概要

- 前回までの消費統計研究会において、2019年全国家計構造調査結果における標準誤差の推定方法として、ブートストラップ法の有効性を確認
- 今回は引き続きブートストラップ法により、格差指標についての誤差推定を行う

前回の試算の主な内容

平成26年全国消費実態調査

- ブートストラップ法の有効性の確認（理論式との比較など）

2019年全国家計構造調査（2014年遡及結果、2019年結果）

- ブートストラップ法による誤差推定（消費支出）

今回の試算の主な内容

2019年全国家計構造調査

- ブートストラップ法による誤差推定
 - 等価可処分所得のジニ係数
 - 相対的貧困率

試算におけるリサンプリングの方法

※各抽出段階における抽出率の計算には平成27年国勢調査結果を使用する。

二人以上の世帯

- 市：層化2段抽出に Bernoulli Bootstrap（以下、BBE）を適用
市別に調査単位区抽出 → 調査世帯抽出
- 町村：層化3段抽出にBBEを適用
都道府県別に町村抽出 → 調査単位区抽出 → 調査世帯抽出

単身世帯

- 市：市別に層化1段抽出のBBEを適用
1世帯のみ抽出の調査単位区が多いため、市別に調査世帯抽出
- 町村：層化2段抽出にBBEを適用
都道府県別に町村抽出 → 調査世帯抽出
- モニター調査：都道府県別に単純なブートストラップ法を適用

留意点

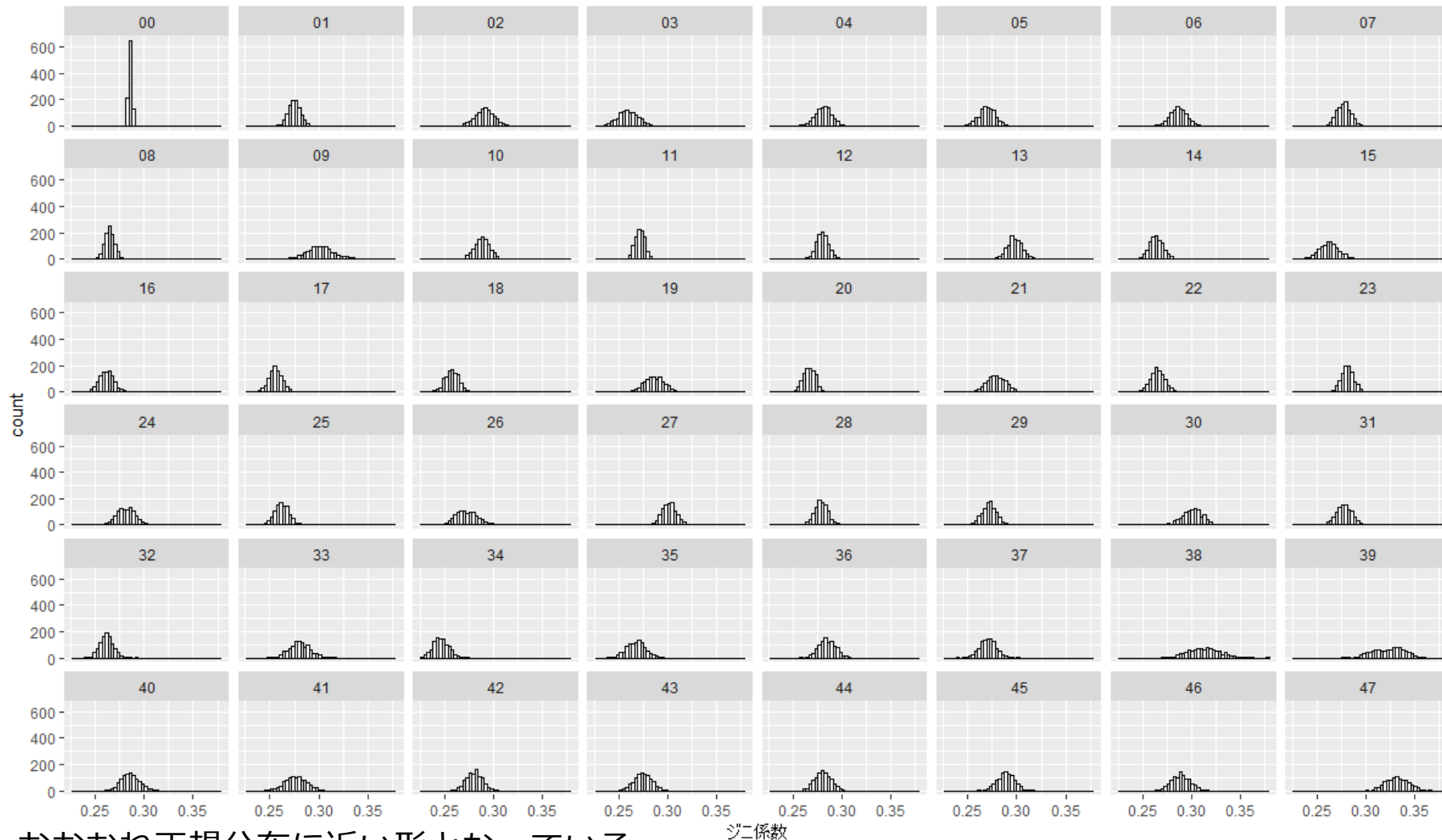
- 基本調査＋特別調査、簡易調査のそれぞれについて独立に抽出する
- 各抽出段階において抽出対象が1つしか存在しない場合、別の区分と統合する

試算結果：等価可処分所得のジニ係数

都道府県別・ブートストラップ平均の分布

※ブートストラップ回数：1000回

可処分所得はOECDの新基準準拠



おおむね正規分布に近い形となっている。

(歪度： $-0.22 \sim 0.34$ 尖度：高知県 (39) で -0.77 、ほかは $-0.37 \sim 0.46$)

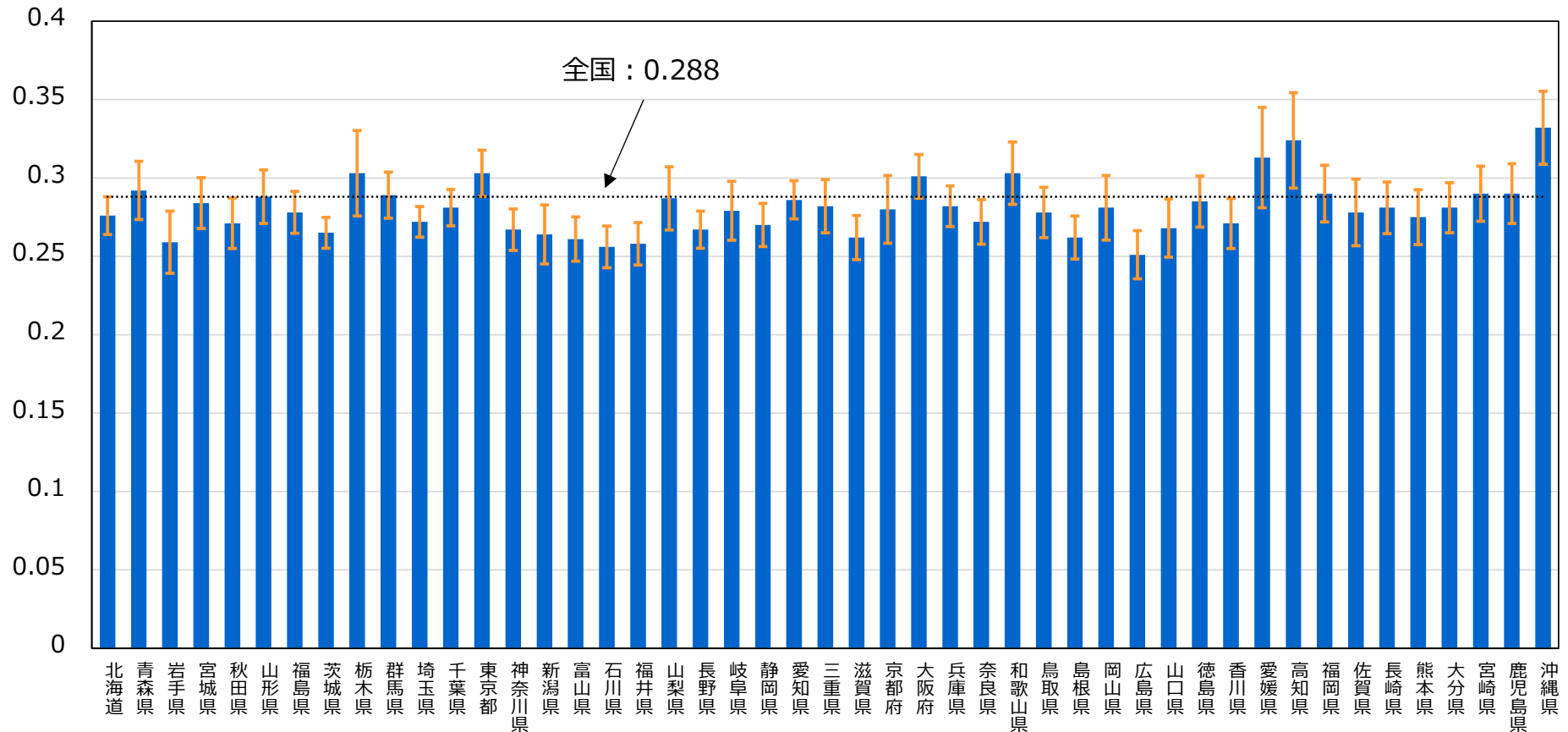
試算結果：等価可処分所得のジニ係数

都道府県別

※ブートストラップ回数：1000回

可処分所得はOECDの新基準準拠

都道府県別ジニ係数



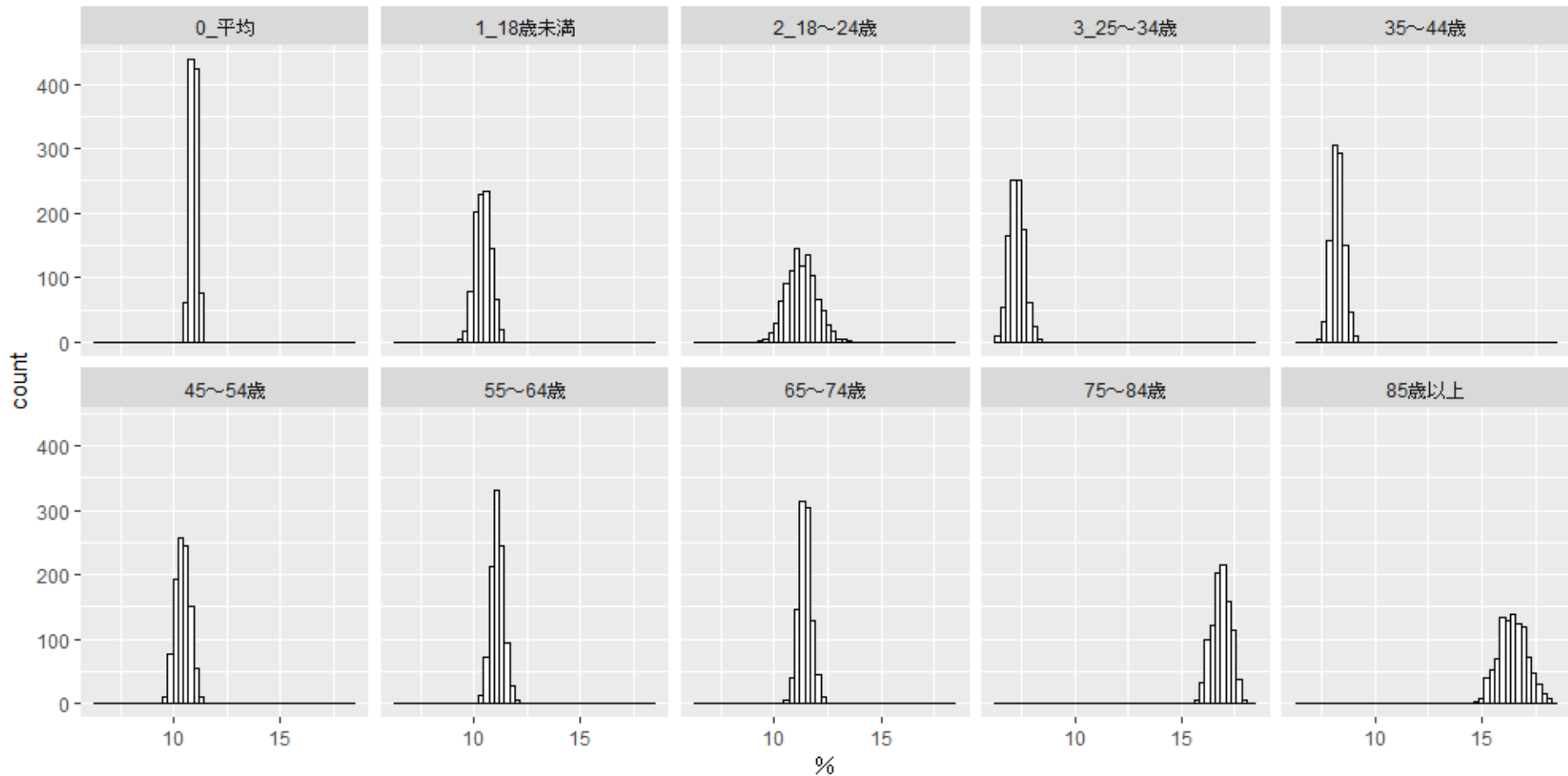
※オレンジの線は95%信頼区間

ジニ係数に対する標準誤差の大きさは 1.8% ~ 5.2%、中央値は 3.0%

試算結果：相対的貧困率

年齢階級別・ブートストラップ平均の分布

※ブートストラップ回数：1000回
可処分所得はOECDの新基準準拠



おおむね正規分布に近い形となっている。

(歪度：-0.06~0.23 尖度：-0.30~0.28)

試算結果：相対的貧困率

年齢階級別

※ブートストラップ回数：1000回
可処分所得はOECDの新基準準拠

| 年齢階級 | 集計人員 | 相対的貧困率 (公表値・%) | ブートストラップ 平均のバイアス (ポイント) | 標準誤差 (ポイント) | 相対的貧困率 に対する標準 誤差の大きさ (%) |
|--------|---------|-------------------|-------------------------------|----------------|-----------------------------------|
| 平均 | 199,489 | 11.2 | ▲ 0.30 | 0.16 | 1.5 |
| 18歳未満 | 32,870 | 10.3 | 0.10 | 0.37 | 3.6 |
| 18～24歳 | 9,434 | 11.8 | ▲ 0.50 | 0.71 | 6.0 |
| 25～34歳 | 15,072 | 7.3 | ▲ 0.06 | 0.35 | 4.8 |
| 35～44歳 | 24,328 | 8.2 | 0.10 | 0.30 | 3.6 |
| 45～54歳 | 27,702 | 10.4 | ▲ 0.01 | 0.34 | 3.3 |
| 55～64歳 | 27,548 | 11.3 | ▲ 0.25 | 0.30 | 2.6 |
| 65～74歳 | 34,152 | 11.9 | ▲ 0.49 | 0.30 | 2.5 |
| 75～84歳 | 21,303 | 18.3 | ▲ 1.48 | 0.45 | 2.4 |
| 85歳以上 | 7,080 | 16.8 | ▲ 0.32 | 0.67 | 4.0 |

相対的貧困率に対する標準誤差の大きさは 2.4% ～ 6.0% (中央値は3.4%)

※ 都道府県別の場合を試算すると、6.3% ～ 15.4% (中央値は 9.9%)

結果のまとめ

- 2019年全国家計構造調査結果のうち格差指標に関する結果について、ブートストラップ法による誤差推定の試算を行った。
- 公表している格差指標（都道府県別のジニ係数、年齢階級別の相対的貧困率）については、高い精度で推定されていることが確認できた。
- 都道府県別の相対的貧困率については、標本誤差が大きく、公的統計としての公表には課題が残ることが確認できた。
- 引き続き、消費、所得、資産等の精度について、研究を進める予定

(参考) Bernoulli Bootstrap の手順

第1段：層 h から抽出する第1抽出区分のサンプルサイズ n_h ，層 h からの抽出率 f_{1h}

第2段：第1抽出区分 hi から抽出する第2抽出区分のサンプルサイズ m_{hi} ，抽出率 f_{2hi}

第3段：第2抽出区分 hij から抽出する標本のサンプルサイズ l_{hij} ，抽出率 f_{3hij}

ステップ1

層 h から $n_h - 1$ 個の第1抽出区分をランダムに復元抽出した集合を \widetilde{PSU}_h とする。層 h に含まれる第1抽出区分 hi ($i = 1, 2, \dots, n_h$)について、以下の確率 p_h

$$p_h = 1 - \frac{1}{2} \frac{1 - f_{1h}}{1 - n_h^{-1}}$$

でそのまま残し、確率 $1 - p_h$ で \widetilde{PSU}_h からランダムに選んだ他の第1抽出区分と入れ替える。

ステップ2

第1抽出区分 hi から $m_{hi} - 1$ 個の第2抽出区分をランダムに復元抽出した集合を \widetilde{SSU}_{hi} とする。ステップ1でそのまま残した区分 hi に含まれる第2抽出区分 hij ($j = 1, 2, \dots, m_{hi}$)について、以下の確率 q_{hi}

$$q_{hi} = 1 - \frac{1}{2} \frac{f_{1h}}{p_h^{-1}} \frac{1 - f_{2hi}}{1 - m_{hi}^{-1}}$$

でそのまま残し、確率 $1 - q_{hi}$ で \widetilde{SSU}_{hi} からランダムに選んだ他の第2抽出区分と入れ替える。

ステップ3

第2抽出区分 hij から $l_{hij} - 1$ 個の標本をランダムに復元抽出した集合を \widetilde{USU}_{hij} とする。区分 hij に含まれる標本 $hijk$ ($k = 1, 2, \dots, l_{hij}$)について、以下の確率 r_{hij}

$$r_{hij} = 1 - \frac{1}{2} \frac{f_{1h}}{p_h^{-1}} \frac{f_{2hi}}{q_{hi}^{-1}} \frac{1 - f_{3hij}}{1 - l_{hij}^{-1}}$$

でそのまま残し、確率 $1 - r_{hij}$ で \widetilde{USU}_{hij} からランダムに選んだ他の標本と入れ替える。