

# 国民生活基礎調査における所得分布の精度について

イシイ フトシ フルヤ ヒロフミ  
石井 太\*<sup>1</sup> 古屋 裕文\*<sup>2</sup>

**目的** 国民生活基礎調査の実施により得られる所得分布に関する情報提供の充実に資する観点から、同調査の所得分布に関する精度について検討を行う。

**方法** 国民生活基礎調査における所得分布に関する誤差情報について、ブートストラップ法などを用い、所得分位値、所得階級別構成割合、ジニ係数の各標準誤差(率)に関する評価を行った。

**結果** 中央値の標準誤差率は大規模年では1.3%、中間年では2.4%となっており、他の分位値とともに、概して平均値よりも誤差率が高かった。所得階級(100万円階級)別構成割合については、1%の幅を許容すれば95%の確率で利用できるものと考えられた。ジニ係数の標準誤差は、平成16年調査では0.0027ポイント、平成15年調査では0.0050ポイントであった。

**結論** 従来、評価が困難であった分位値などに関する精度の定量的評価がブートストラップ法を用いることにより可能となり、対前年比較や結果の信頼性など、所得分布に関する各種指標や統計表の見方に関する豊富な情報を提供することができる。

**キーワード** 所得分布、分位値、標準誤差、ブートストラップ法、ジニ係数

## I はじめに

年金制度をはじめとする所得保障にとって、所得統計は制度検討の基礎となる重要な統計である。特に、世帯を単位として様々な制度構築が行われてきた厚生行政の分野では、世帯を単位とする所得統計はあらゆる制度統計の前提とみることもでき、その重要性・影響力も大きいものとなっている。このような情報を提供するのが、厚生労働省統計情報部において実施されている国民生活基礎調査(指定統計第116号)である。同調査では、3年に1回の大規模調査では約2万5千世帯(平成16年調査)、中間年の調査では約7千世帯(平成15年調査)に対して所得調査を行っており、わが国の世帯ベースでの所得情報を継続的に提供している。

ところで、この国民生活基礎調査は全数調査ではなく、調査対象から一部の客体を標本として無作為に抽出する標本調査である。標本調査では標本抽出に起因する「標本誤差」というものが存在するため、調査の実施に当たっては、調査結果の精度を示す「標本誤差」を一定のレベルに管理し、客体数や抽出方法などの策定を行う「標本設計」が必須のものとなる<sup>1)</sup>。一般的に、調査の結果表は対象となる指標の総和や平均を表示することから、標本誤差も総和や平均に関するものを利用・表章することが多い。国民生活基礎調査でも所得票については1世帯当たり平均所得額の標準誤差率が報告書に掲載されている。

一方で、所得統計は平均値のみならず、中央値などの分位値や分布それ自体も重要な指標で

\* 1 厚生労働省大臣官房統計情報部企画課審査解析室長補佐(現国立社会保障・人口問題研究所企画部第四室長)

\* 2 同部企画課審査解析室総合解析係

あり、調査結果としてもよく利用されている。「統計行政の新たな展開方向」(平成15年6月各府省統計主管部局長等会議申合せ)の中でも、「各府省は、利用者の利便性を確保する観点から、これまでの分析事例を参考に平均値、合計値だけでなく分布に関する各種統計量等調査の特性に応じた一層の分析を行い、これを提供する」とされるなど、分布情報の提供は重要な課題といえる。さらに、標本調査では標本誤差情報の提供も重要であり、「統計行政の新たな展開方向」でも、指定統計については達成誤差などの誤差情報を提供していくこととされたほか、既に情報提供している統計調査についても「その内容の充実を図ることとし、承認統計や届出統計についても指定統計に準じて情報提供を図ること」とされており、分布情報の充実にあわせてその誤差情報の提供が求められてきている。しかしながら、国民生活基礎調査の場合、報告書レベルでは所得分布に関する誤差情報の提供が行われていない。

こうした現状を踏まえ、国民生活基礎調査における所得分布に関する誤差情報の提供を目的とし、ブートストラップ法を応用する研究<sup>2)3)</sup>や、所得分位値に関する標準誤差を推定する研究<sup>4)5)</sup>などが行われてきている。

本稿では、これらの研究の延長として、直近の大規模調査である平成16年調査で所得分位値の標準誤差を評価し、中間年である平成15年調査と比較した。また、所得階級(100万円階級)別構成割合の標準誤差、所得格差を表す指標であるジニ係数の標準誤差についても定量的評価を行い、国民生活基礎調査の所得分布の精度に関する包括的な検討を行った。

## II 方 法

国民生活基礎調査所得票の1世帯当たり所得額について、調査データを用いて、平均値、分位値(四分位値・十分位値)、所得階級(100万円階級)別構成割合、ジニ係数に関する標本誤差推定を行う。平均値と構成割合については、理論式による標準誤差(率)推定の評価を行い、

特に平均値については、ブートストラップ法としてBWR(With Replacement Bootstrap)法に基づく推定値<sup>6)</sup>も算定して比較を行った。分位値(四分位値・十分位値)とジニ係数については、ブートストラップ法のみによる標準誤差(率)の推定を行い、特に分位値については、ブートストラップ標本を用いた信頼区間の構成も行った。具体的な推定方法(大規模年の場合)を以下に述べる。

平均所得の推定式は、

$$\hat{R} = \frac{\sum_k \left( \frac{N_k M_k}{n_k m_k} \sum_j X_{kj} \right)}{\sum_k \left( \frac{N_k M_k}{n_k m_k} \sum_j Y_{kj} \right)} = \frac{\sum_k \left( L_k \frac{1}{m_k} \sum_j X_{kj} \right)}{\sum_k \left( L_k \frac{1}{m_k} \sum_j Y_{kj} \right)}$$

で与えられる。ここで、

$\hat{R}$  : 推計値

$N_k$  :  $k$  県の国勢調査地区数(後置番号1)

(都道府県・政令指定都市を「県」と略記)

$n_k$  :  $k$  県の世帯票調査地区数(後置番号1)

$M_k$  :  $k$  県の  $n_k$  個の調査地区から設定された単  
位区数

$m_k$  :  $k$  県の調査単位区数

$X_{kj}$  :  $k$  県  $j$  単位区の世界の総所得

$Y_{kj}$  :  $k$  県  $j$  単位区の世界の総数

$$L_k = \frac{N_k M_k}{n_k}$$

である。

理論式に基づく推定値の分散の評価式は以下のとおりである。

$$V(\hat{R}) = \frac{1}{Y^2} \sum_k \left( \frac{L_k}{L} \right)^2 \left( \frac{1}{m_k} - \frac{1}{L_k} \right) \times (V_k(X) - 2\hat{R}COV_k(X, Y) + \hat{R}^2 V_k(Y))$$

ただし、

$$L = \sum_k L_k$$

$$V_k(X) = \frac{1}{m_k - 1} \sum_j (X_{kj} - \bar{X}_k)^2$$

$$V_k(Y) = \frac{1}{m_k - 1} \sum_j (Y_{kj} - \bar{Y}_k)^2$$

$$COV_k(X, Y) = \frac{1}{m_k - 1} \sum_j (X_{kj} - \bar{X}_k)(Y_{kj} - \bar{Y}_k)$$

$$\bar{X}_k = \frac{1}{m_k} \sum_j X_{kj}$$

$$\bar{Y}_k = \frac{1}{m_k} \sum_j Y_{kj}$$

$$\bar{X} = \frac{1}{L} \sum_k L_k \frac{1}{m_k} \sum_j X_{kj}$$

$$\bar{Y} = \frac{1}{L} \sum_k L_k \frac{1}{m_k} \sum_j Y_{kj}$$

であり、このとき、標準誤差の推定値は $\sqrt{V(\hat{R})}$ 、標準誤差率の推定値は $\frac{\sqrt{V(\hat{R})}}{\hat{R}}$ で表される。

これに対し、BWR法による誤差評価は以下の手順のとおりである。最初に各層における抽出率を $f_k$ とし、各層ごとのブートストラップ標本のサイズを、

$$s_k = \frac{m_k - 1}{1 - f_k}$$

により決定する。次に、各層において、大きさ $s_k$ のブートストラップ標本(右肩に\*を付けたものはブートストラップ標本を表す)

$$\{(X_{kj}^*, Y_{kj}^*) : j = 1, \dots, s_k\}$$

を、標本

$$\{(X_{kj}, Y_{kj}) : j = 1, \dots, m_k\}$$

から復元抽出する。これをB回繰り返す。b回目の復元抽出によるブートストラップ標本における平均所得については、

$$\bar{X}_k^{(b)} = L_k \cdot \frac{1}{s_k} \sum_j X_{kj}^*, \quad \bar{Y}_k^{(b)} = L_k \cdot \frac{1}{s_k} \sum_j Y_{kj}^*$$

表1 標準誤差(率)推定結果

	平成16年調査(大規模調査)			平成15年調査(中間年調査)		
	推計値(万円)	標準誤差(万円)	標準誤差率(%)	推計値(万円)	標準誤差(万円)	標準誤差率(%)
平均値(平均所得額)	579.7	5.3	0.9	589.3	9.9	1.7
第1四分位値	263	4.5	1.7	255	8.6	3.3
第2四分位値(中央値)	476	6.0	1.3	476	11.3	2.4
第3四分位値	780	7.4	0.9	791	12.5	1.6
第1十分位値	133	2.6	2.0	129	5.8	4.5
第2十分位値	219	4.1	1.9	213	8.2	3.8
第3十分位値	303	3.6	1.2	303	7.9	2.6
第4十分位値	390	4.8	1.2	390	8.8	2.3
第5十分位値	476	6.0	1.3	476	11.3	2.4
第6十分位値	582	7.2	1.2	582	12.6	2.2
第7十分位値	700	6.2	0.9	708	12.8	1.8
第8十分位値	862	8.0	0.9	880	16.2	1.8
第9十分位値	1 117	10.2	0.9	1 153	18.5	1.6

を用いて、

$$R^{(b)} = \frac{\sum_k \bar{X}_k^{(b)}}{\sum_k \bar{Y}_k^{(b)}} = \frac{\sum_k L_k \frac{1}{s_k} \sum_j X_{kj}^*}{\sum_k L_k \frac{1}{s_k} \sum_j Y_{kj}^*}$$

を計算し、このブートストラップ標本に対応する推定量を作成する。また、分位値については、各世帯の所得がそれぞれ $\frac{L_k}{s_k}$ のウェイトをもっていると考え、累積分布関数の各分位点に対応する世帯の所得を $R^{(b)}$ とし、このブートストラップ標本に対応する推定量とする。ジニ係数も同様に推定量が構成可能である。その後、これらのブートストラップ標本に対応するB個(B=100,000)の推定量から、Rの推定量

$$\hat{R} = \frac{1}{B} \sum_b R^{(b)}$$

を推定し、分散の推定量を、

$$V(\hat{R}) = \frac{1}{B-1} (\sum_b R^{(b)2} - B \cdot \hat{R}^2)$$

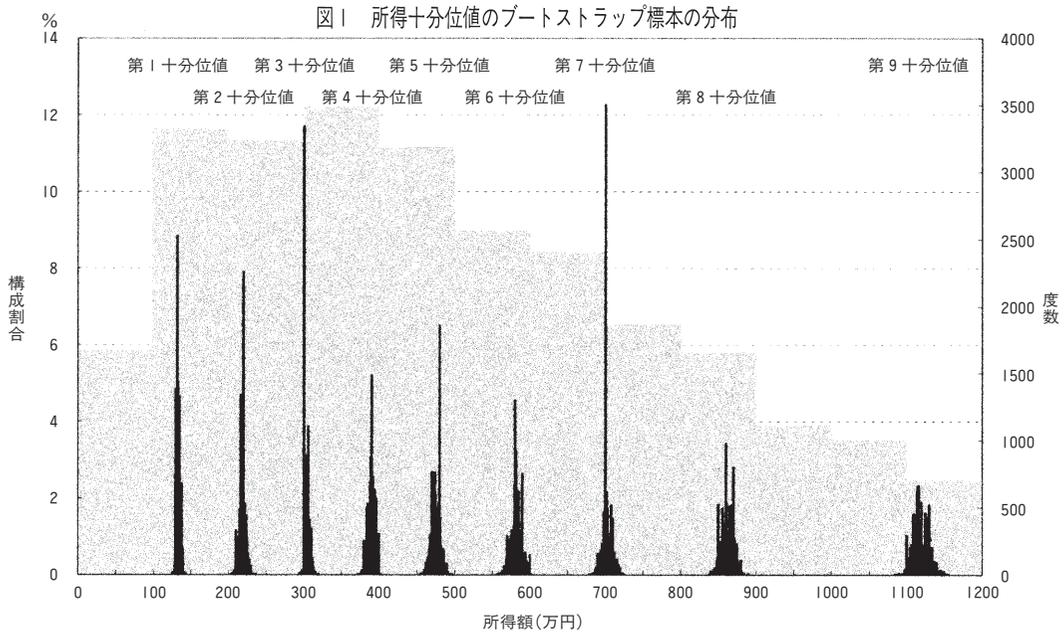
により推定する。また、ブートストラップ標本の分布を用いることにより、推定量の信頼区間を構成することができる。

### III 結果と考察

#### (1) 標準誤差(率)の推定

平成15年、16年の国民生活基礎調査に基づく標準誤差(率)の推定結果を表1に示す。報告書にも掲載されている平均所得額については、理論式に基づく結果を示してある。この結果についてはBWR法に基づいて標準誤差(率)を推定してもほぼ同一の値となることから、BWR法による標準誤差推定の妥当性を裏付けているといえる。

平均値の標準誤差率は、大規模調査で0.9%、中間年調査で1.7%の水準となっており、標準誤差率でみた場合、大規模年は中間年の約2倍の精度が確保されていることがわかる。一方、分位値に関する標準誤差(率)はBWR法に基づく推定結果であり、従来の方法である理論式からは得られなかつ



た結果である。これによると、例えば中央値の標準誤差率は大規模年で1.3%、中間年で2.4%となっており、平均値の標準誤差率0.9%、1.7%よりもやや高い推定結果となっている。他の分位値についても概して平均値よりも誤差率が高いものとなっており、分位値を利用の際は平均値に比べて注意が必要であることがわかる。また、所得が低い方の分位値では標準誤差率が高いのに対し、所得が高い方の分位値に向かうにつれて標準誤差率は低くなる傾向がみられる。ただし、所得が高い方の分位値はそもその推計値自体が大きいため、標準誤差率は小さくても標準誤差の絶対値はそれなりの大きさがあることに留意すべきであろう。

(2) 信頼区間の推定

標準誤差(率)は推定量の精度を表す基本的な量であるが、推定量に中心極限定理が適用できることを前提とし、推定量が正規分布すると考えて信頼区間を推定するためにも用いられる。例えば、平成16年調査の平均所得の場合、推定値が579.7万円であり、標準誤差が5.3万円であることから、推定値の両側に1.96倍の幅を取った [569.3, 590.1] を95%信頼区間とするものである。これに対し、一般に分位値では、中心

表2 所得十分位値の95%信頼区間

(単位 万円)

十分位値	95%信頼区間	
	下限	上限
第1十分位値	133	139
第2十分位値	219	228
第3十分位値	303	311
第4十分位値	390	400
第5十分位値	476	487
第6十分位値	582	598
第7十分位値	700	717
第8十分位値	862	878
第9十分位値	1 117	1 140

極限定理により推定量が正規分布することが必ずしも保証されていないため、ブートストラップ標本から信頼区間を推定する方法が有力なものとなる。これにはいくつかの方法が研究されているが、ここでは最も簡単なものとしてブートストラップ標本の分布から直接信頼区間を構成するパーセンタイル法により信頼区間を推定することとした。ブートストラップ標本の分布自体を図1に示す。

この図では100,000個のブートストラップ標本に基づき所得十分位値の分布を作成したが、これによれば、実際に所得十分位値では、ブートストラップ標本の分布の形状は必ずしも正規分

布に近いものとはなっていないことがわかる。特に、第3十分位値や第7十分位値では、分位値自体が303万円、700万円と、100万円刻みで切りのよい値か、あるいはそれに極めて近い値となっており、調査票でも切りのよい値を記入する者が比較的多いことなどからその点に標本が集中する分布となっている。これを用いて、所得分位値の信頼区間を作成したものが表2である。

これからわかるように、所得分位値の95%信頼区間に関しては、正規近似を前提として単純に標準誤差の1.96倍を推定値の左右に取ったものとはやや様相が異なるものとなっている。特に、ブートストラップ標本の分布でやや特異値を含んでいた第3十分位値や第7十分位値では信頼区間の左右の幅がかなり異なっていることが読み取れる。

### (3) 所得階級別構成割合の精度

ここまで、主に所得分位値の標準誤差(率)と推定量の信頼区間を評価することにより所得分布の精度について論じた。しかしながら、所得分布を表す別の方法として所得階級別の構成割合をみる方法があり、実際に国民生活基礎調査においても所得階級(100万円階級)別構成割合による所得分布が表章されている。しかしながら、これまで報告書においてその標準誤差率は掲載されていない。そこで、所得分布の精度を論じる観点から、関連として所得階級(100万円階級)別構成割合の標準誤差(率)を評価してみる。

平成16年調査による誤差推定結果を表したものが表3である。これによれば、所得階級(100万円階級)別構成割合の標準誤差が最も大きくなっているのは100~200万円階級の構成割合11.6%に対する0.31%ポイントであり、これに対応する95%信頼区間の幅は1.2%ポイントとなる。すなわち、所得階級(100万円階級)別構成割合の調査結果は、全数調査を行って得られるはずの真の構成割合とのブレについて、1%ポイントの幅を許容すれば95%の確率で利用できるといえよう。また、特に高所得階級では構成

表3 所得階級別構成割合とその標準誤差(率)

(単位 %)

所得階級	構成割合	標準誤差	標準誤差率
100万円未満	5.9	0.20	3.4
100~200	11.6	0.31	2.7
200~300	11.3	0.24	2.2
300~400	12.2	0.27	2.2
400~500	11.2	0.26	2.3
500~600	9.0	0.22	2.5
600~700	8.4	0.23	2.7
700~800	6.5	0.19	2.8
800~900	5.8	0.19	3.3
900~1000	3.9	0.16	4.1
1000~1100	3.5	0.15	4.3
1100~1200	2.5	0.11	4.6
1200~1300	2.0	0.10	5.2
1300~1400	1.4	0.09	6.4
1400~1500	1.1	0.09	7.5
1500~1600	0.9	0.07	8.2
1600~1700	0.6	0.05	9.8
1700~1800	0.4	0.04	11.6
1800~1900	0.4	0.05	12.5
1900~2000	0.2	0.03	14.4
2000万円以上	1.2	0.09	7.3

割合の絶対値が小さくなるため標準誤差率が大きくなってしまいが、例えば、2000万円以上の階級では構成割合が1.2%であるのに対して標準誤差は0.09%ポイントとなっており、95%信頼区間は[1.1, 1.4]となる。したがって、誤差率はやや大きいものの、2000万円以上所得がある世帯の割合はおおむね1%代前半にはあるといえよう。

### (4) ジニ係数の精度

さて、所得分布を考える別の視点として、所得格差の測定がある。所得格差を表す指標にはSCV, MLDなどいくつかのものが考えられているが、このうち最も代表的に用いられてきているのがジニ係数である。国民生活基礎調査の世帯所得額について、BWR法に基づいて推定したジニ係数の標準誤差(率)を表4に示す。

ジニ係数は平成16年調査では0.3882と前年の平成15年調査に比べ0.0104ポイント低下し、前年比2.6%の減少となっている。この結果から所得格差は減少しているといえるのであろうか。標準誤差は平成16年調査では0.0027ポイント、平成15年調査では0.0050ポイントとなっており、正規近似が成立すると仮定すると、95%信頼区間の幅は平成16年調査では0.0108、平成15年調

査では0.0196となり、これらは対前年増減よりも大きい。特に平成15年調査は中間年調査であり、信頼区間も大規模年調査に比べ約2倍の大きさとなっていることから、所得格差の変動に関しては、より長期的な動向を観察するなどによる慎重な検討が必要であるといえよう。

なお、実際の調査データに基づくジニ係数の精度評価を行ったものとして小林による先行研究<sup>7)</sup>があり、昭和56年・平成5年所得再分配調査（国民生活基礎調査所得票の中間年と同一客体）のデータを用い、ブートストラップ法によりジニ係数の標準誤差を推定している。これによれば、例えば平成5年調査の場合、再分配後所得で0.0035との標準誤差推定結果を得ており、この結果は、筆者らが得た中間年調査（平成15年調査）の標準誤差推定結果0.0050におおむね近いものの、やや小さめの値となっている。この差については、年次や対象とする所得の違いのほか、小林は世帯を抽出単位とした無作為抽出が行われたとしてブートストラップ法を適用していると考えられるのに対し、実際の標本抽出では国勢調査地区を分割した単位区を抽出単位とする集落抽出法が取られており、筆者らはこの標本抽出法を考慮したブートストラップ法を適用して評価を行っていることなどに起因すると考えられる。

#### IV おわりに

所得分布は所得調査の利用者にとって重要な統計表であるが、その利用に当たっては標準誤差など精度に関する定量的な評価が必須のものとなる。従来、その定量的評価は難しい面があり、行われてこなかったが、ブートストラップ法を用いることにより定量的評価が可能であり、その結果を用いれば、対前年比較や結果の信頼性など、所得分布に関する各種指標や統計表の見方に関する豊富な情報を提供することが可能であることがわかった。

表4 ジニ係数の標準誤差(率)と95%信頼区間

	平成16年調査（大規模調査）			平成15年調査（中間年調査）		
	推計値	標準誤差	標準誤差率	推計値	標準誤差	標準誤差率
ジニ係数	0.3882	0.0027	0.7%	0.3986	0.0050	1.3%
対前年増減	△0.0104	△2.6%				
95%信頼区間(下限)	0.3828			0.3888		
95%信頼区間(上限)	0.3936			0.4084		
95%信頼区間の幅	0.0108			0.0196		

国民生活基礎調査の所得分布に関しては、世帯票などの補助情報を活用して推定精度を高める方法がいくつか提案されており、現在、このような方法を応用する研究<sup>8)</sup>が行われているところであるが、所得分布に関するより高度な結果の提供に資する観点からも、今後、このような分野におけるさらなる研究が必要である。

#### 謝辞

本研究に関連し、貴重なご助言、ご協力を賜った早稲田大学の西郷浩教授、群馬大学の樋田勉助教授に対し、感謝の意を表します。

#### 文 献

- 1) 村山令二, 鈴木健二, 石井太, 他. よくわかる標本調査法. 東京: 厚生統計協会, 2004; 32-139.
- 2) 馬場康雄. 官庁統計におけるブートストラップ法の利用. 第65回統計学会チュートリアルセミナー. 1997.
- 3) 馬場康雄, 土屋隆裕, 中村好宏, 他. 比推定における標準誤差のブートストラップ推定の試み. 第64回日本統計学会講演報告集. 1996; 240-1.
- 4) 石井太. 国民生活基礎調査標本誤差推定へのリサンプリング法の応用. 労働統計調査月報 2004; 56(1): 12-8.
- 5) 石井太, 鈴木健二, 西郷浩. 国民生活基礎調査標本誤差推定へのリサンプリング法の応用. 2003年度統計関連学会連合大会講演報告集. 2003; 491-2.
- 6) Shao J, Tu D. The Jackknife and Bootstrap. Springer 1995.
- 7) 小林正人. 平成9年度厚生科学研究費補助金(厚生行政科学研究事業) 所得再分配の評価手法に関する研究報告(代表 石井経夫). 1998; 104-16.
- 8) 石井太, 關雅夫, 西郷浩, 他. 二相抽出法を利用した国民生活基礎調査所得分布推定の検討. 2004年度統計関連学会連合大会講演報告集. 2004; 54-5.