

国民生活基礎調査における無回答データ等の 影響を考慮した調査設計に関する研究

ニッダ
新田

イナ
功*

目的 国民生活基礎調査による世帯数の推計値に無回答データが及ぼす影響を補正する方法について検討することを目的とした。

方法と結果 性別・年齢階級別人口について国民生活基礎調査の推計値と他の調査の推計値とを比較すると乖離がみられる。その原因は、国民生活基礎調査における無回答（面接不能等）世帯の増加にあると考えられ、国勢調査等との比較を行うと、国民生活基礎調査における無回答世帯が65歳未満の単独世帯、特に20歳代、30歳代の単独世帯に多いことが明らかになった。無回答世帯に起因する非標本誤差の影響を除去する方法として加重法と補完法があるが、国民生活基礎調査においては調査不能世帯の属性に関する情報が得られることから、この情報に基づいて、推計の際に用いる加重を調整する方法を検討した。数種類の加重を比較検討のうえ、世帯を、①65歳未満の単独世帯、②65歳以上の単独世帯、③2人以上世帯に区分し、それぞれのグループの回収客体数に国勢調査に基づいて決定した加重を乗じて推計値を求める方法を採用した。その結果、世帯を3種類に区分して異なる大きさの加重を乗じるこの補正方法により、国民生活基礎調査の世帯数の全国推計値の改善がみられた。

結論 国民生活基礎調査による世帯総数、単独世帯数等の推計値を実態に近づけるという点において、本研究の補正方法は有効である。しかし、この補正方法を利用する場合には、従来の方法による推計結果との接合方法等の検討が必要である。

キーワード 国民生活基礎調査、無回答、補正、加重法、世帯票、世帯数の推計

I 緒 言

わが国の代表的な標本調査の1つである国民生活基礎調査（指定統計第116号）においても、プライバシー意識の高まりなどを背景として、回収率が低下傾向にある。回収率の低下は調査の信頼性の低下をもたらし可能性がある。筆者を代表者とする研究会は、国民生活基礎調査の世帯票を分析対象として、無回答データ（全項目無回答標本と調査不能標本の合計）が同調査の世帯数の推計値や分布にもたらす影響の評価を行い、この影響を補正する方法について研究

を行った¹⁾。

標本調査における誤差は標本誤差と非標本誤差に大別できるが、無回答データがもたらす誤差は非標本誤差に属する²⁾³⁾。一般に、無回答データによる誤差の要因として次の2つが指摘されている。

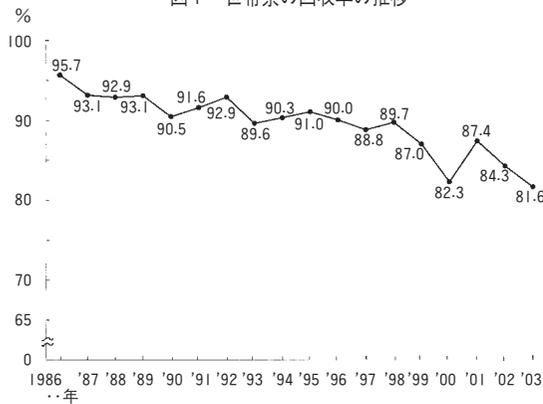
第1は、属性分布の歪みによる誤差であり、第2は、同じ属性の中でも回収標本と無回答標本とで回答分布が異なることに起因する誤差である。この2つの要因による誤差のうち、前者によるものについては補正が可能であり、他方、後者の要因による誤差については補正が不可能であるとされる⁴⁾。

前者の属性分布の歪みを補正する方法として、

* 明治大学政治経済学部教授

加重法 (weighting) と補完法 (imputation) がある。加重法は、無回答の発生に応じて、推定値を計算するときの加重を変更する方法であり、補完法は、無回答を何らかの方法で予測し、その予測値を擬似的な観察値とする方法である⁵⁾。本研究で分析の対象とした世帯票については、回答が得られなくても、無回答の理由、性別、年齢別、地点などの補助情報が得られることから、加重法を対象として検討を行った。

図1 世帯票の回収率の推移



資料 厚生労働省「国民生活基礎調査」
注 1986、89、92、95、98、2001年は大規模調査である。

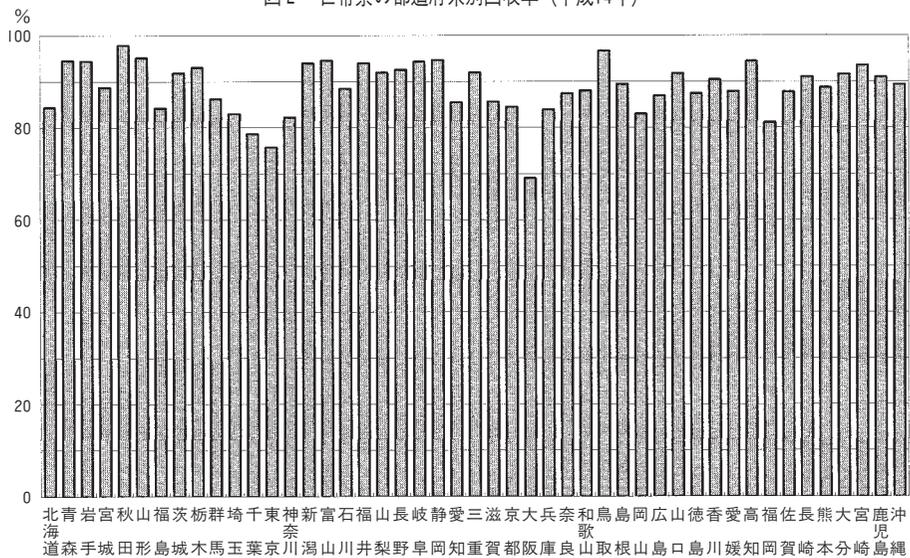
II 研究方法

(1) 国民生活基礎調査における無回答データの発生状況

国民生活基礎調査における回収率は低下傾向にある。回収率は平成8(1996)年に90%であったが、平成15(2003)年には81.6%へと大幅に低下している。3年ごとに行われる大規模調査とその中間年に行われる小規模調査を比較すると、大規模調査の方が1~3ポイント回収率が高いものの、その大規模調査でも回収率は低下している(図1)。

都道府県別に回収率を調べると、平成14(2002)年調査では回収率90%以上が22県、80%台が22県であった。回収率が80%未満は、大阪府(69%)、東京都(75.7%)、千葉県(78.6%)であった(図2)。調査客体についての名簿を用いて無回答の属性を調べると、単独世帯に無回答が多く、大都市を含んでいる都道府県では、単独世帯の無回答の割合は6割に及んでいることが明らかになった¹⁾。

図2 世帯票の都道府県別回収率(平成14年)

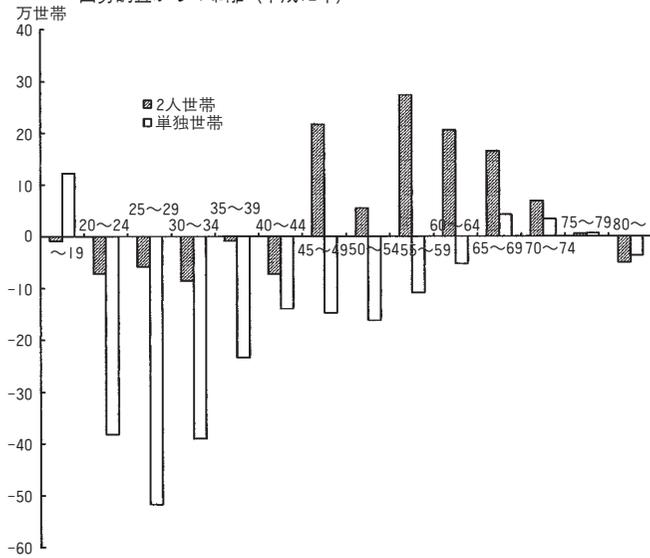


資料 厚生労働省「平成14年 国民生活基礎調査」

(2) 国勢調査との世帯数の比較

名簿の検討により単独世帯に無回答が多いことが特定できたが、無回答が多いのはどの年齢階級の単独世帯であるのかを確かめるには他調査との比較が必要である。世帯類型別に世帯の調査を定期的に行っているのは国勢調査だけなので、平成12(2000)年の国勢調査の結果と同年の国民生活基礎調査の世帯推計とを比較⁶⁾した。その結果、単独世帯数では、19歳以下と65~79歳で国民生活基礎調査の推計値が国勢調査の結果よりも多少多くなっているが、他のすべての年齢階級で少なくなっていること、特に20歳代では89万9000世帯も少なくなっていることが明らかになった。また、2人以上世帯では、世帯主の年齢が45~79歳では国民生活基礎調査の推計値の方が国勢調査の結果よりも多く、他の年齢では少ないことも判明した(図3)。

図3 世帯主の年齢階級別・世帯属性別にみた国民生活基礎調査の推計世帯数の国勢調査からの乖離(平成12年)



資料 厚生労働省「平成12年国民生活基礎調査」、総務省「平成12年国勢調査」

注 1) 国民生活基礎調査には外国人世帯は含まない。

2) グラフは、世帯属性・年齢階級別にみた国勢調査の世帯数と国民生活基礎調査の推計世帯数の差(国勢調査の世帯数-国民生活基礎調査の推計世帯数)を表す。

(3) 無回答世帯の誤差を補正する方法

(1)と(2)で得られた結果に基づいて、本研究では、無回答バイアスの要因は世帯人員と年齢であると考え、無回答世帯の補正を、①65歳未満の単独世帯、②65歳以上の単独世帯、③2人以上世帯、の3区分で行うことにした。その方法は、以下のとおりである。

いま、平成12年の国民生活基礎調査における都道府県別、3区分別の回答世帯数を x_{t1} , x_{t2} , x_{t3} とし、都道府県別の無回答世帯数を y_t とする。さらに、平成12年国勢調査における都道府県別、3区分別の世帯数を X_{t1} , X_{t2} , X_{t3} とする。このとき、3区分別の無回答世帯数の推計値 y'_{t1} , y'_{t2} , y'_{t3} を、以下の式を満たすように求める。

$$\begin{cases} y'_{t1} + y'_{t2} + y'_{t3} = y_t \\ (x_{t1} + y'_{t1}) : (x_{t2} + y'_{t2}) : (x_{t3} + y'_{t3}) = X_{t1} : X_{t2} : X_{t3} \end{cases}$$

つまり、平成12年国民生活基礎調査における(無回答世帯も含めた)3区分別の世帯数の比率は、

国勢調査のそれと一致しているものとする。そして、無回答世帯の中での3区分の比率 $y'_{t1} : y'_{t2} : y'_{t3}$ は、平成12年から現在まで変化しないと仮定する。たとえば、平成14年調査における都道府県別の無回答世帯数を $y_t^{(14)}$ とすれば、それを比率 $y'_{t1} : y'_{t2} : y'_{t3}$ で按分することによって3区分別の無回答世帯 $y'_{t1}^{(14)}$, $y'_{t2}^{(14)}$, $y'_{t3}^{(14)}$ を推定することができる。

世帯数についての最終的な補正値は、推計値に次式を乗じることによって得られる。

$$\frac{x_{t1}^{(14)} + y'_{t1}^{(14)}}{x_{t1}^{(14)}}, \frac{x_{t2}^{(14)} + y'_{t2}^{(14)}}{x_{t2}^{(14)}}, \frac{x_{t3}^{(14)} + y'_{t3}^{(14)}}{x_{t3}^{(14)}}$$

なお、 y'_{t1} , y'_{t2} , y'_{t3} のいずれかが負の値になる場合にはその値はゼロとし、負になった部分は別の部分に按分することにした。

III 研究結果

平成12年の国勢調査と同年の国民生活基礎調査から算出した世帯人員別の無回答世帯に対する加重を用いて、平成14年の国民生活基礎調査のデータに対して補正を行った。

表1 無回答世帯の世帯属性別割合

(単位 %)

	単独世帯		2人以上世帯		単独世帯		2人以上世帯		単独世帯		2人以上世帯
	65歳未満	65歳以上			65歳未満	65歳以上			65歳未満	65歳以上	
北海道	35.0	22.1	42.9	石川	0.0	19.1	80.9	岡山	13.4	0.3	86.4
青森	17.3	2.8	79.9	福島	0.0	0.0	100.0	広島	24.8	8.3	66.9
岩手	100.0	0.0	0.0	山梨	0.0	0.0	100.0	山口	73.1	26.9	0.0
宮城	19.4	0.0	80.6	長野	100.0	0.0	0.0	徳島	100.0	0.0	0.0
秋田	9.5	2.1	88.4	岐阜	38.7	11.4	49.9	香川	54.6	9.0	36.5
山形	100.0	0.0	0.0	静岡	84.9	0.0	15.1	愛媛	91.6	0.0	8.4
福島	82.3	0.0	17.7	愛知	62.1	0.0	37.9	高知	57.2	0.0	42.8
茨城	58.1	0.0	41.9	三重	67.9	0.0	32.1	福岡	21.4	11.4	67.2
栃木	58.7	1.5	39.8	滋賀	63.0	4.4	32.6	佐賀	84.4	15.6	0.0
群馬	0.0	22.8	77.2	京都	0.0	0.0	100.0	長崎	32.2	0.0	67.8
埼玉	48.5	1.5	50.0	大阪	52.0	6.0	42.0	熊本	32.4	0.0	67.6
千葉	53.3	5.1	41.6	兵庫	50.1	8.0	41.9	大分	83.2	0.0	16.8
東京	45.7	6.2	48.1	奈良	50.5	17.1	32.4	宮崎	96.8	3.2	0.0
神奈川	43.7	5.3	51.1	和歌山	95.1	5.0	0.0	鹿児島	82.1	13.1	4.8
新潟	1.6	29.9	68.6	鳥取	86.3	0.0	13.7	沖縄	0.0	2.8	97.2
富山	95.4	4.6	0.0	島根	100.0	0.0	0.0				

資料 厚生労働省「平成12年国民生活基礎調査」、総務省「平成12年国勢調査」

表2 補正前と補正後の単独世帯と2人以上世帯の構成比

	補正前(%)	補正後(%)
総数	100.0	100.0
単独世帯(65歳未満)	16.0	20.9
単独世帯(65歳以上)	7.5	7.3
2人以上世帯	76.5	71.8

資料 厚生労働省「平成14年国民生活基礎調査」

表3 補正前と補正後の世帯構造別世帯数

	推計値(千世帯)		構成比(%)	
	補正前	補正後	補正前	補正後
総数	46 004	47 992	100.0	100.0
住み込み・寄宿舎等の単独世帯	1 044	1 398	2.3	2.9
その他の単独世帯	9 756	12 144	21.2	25.3
夫婦のみの世帯	9 887	9 707	21.5	20.2
夫婦と未婚の子のみの世帯	14 954	14 710	32.5	30.7
ひとり親と未婚の子のみの世帯	2 841	2 803	6.2	5.8
三世帯世帯	4 603	4 405	10.0	9.2
その他の世帯	2 919	2 825	6.3	5.9

資料 厚生労働省「平成14年国民生活基礎調査」

表1に示したのは、都道府県別に計算した65歳未満の単独世帯、65歳以上の単独世帯、2人以上世帯の無回答に対する加重である。国民生活基礎調査では、都道府県別、指定都市別²⁾に推計が行われているので、本研究でも都道府県別、指定都市別の補正を試みたが、指定都市において y'_{t1} 、 y'_{t2} 、 y'_{t3} のいずれかが負の値になってしまう状況が多発したため、都道府県別の補正のみを行った。しかし、都道府県レベルでも y'_{t1} 、 y'_{t2} 、 y'_{t3} のいずれかが負の値になったため、それをゼロとするケースが多かった。

表2に示したのは、全世帯を65歳未満の単独世帯、65歳以上の単独世帯、2人以上世帯に区分した場合の、補正前と補正後の全世帯に占める構成比である。この表から、平成14年の単独世帯の構成比が補正前よりも4.7ポイント増加したことが明らかであり、国民生活基礎調査の世帯推計値を国勢調査の世帯数の構成比に近づける効果があったことを示している。特に65歳未満の単独世帯の構成比を補正する効果が著しい

ことが明らかである。

表3に示したのは、平成14年のデータに関する世帯構造別の補正前と補正後の推計世帯数と構成比である。これで見るとおり、補正を行った結果、世帯総数は補正前に比べて198万8000世帯増加した。

IV 考察

無回答を加重に反映させる本研究の補正方法は、世帯数の推計において所期の成果を収めることができたといえる。しかし、この補正方法には次の制約がある。第1に、平成12年国勢調査における都道府県別、3区分別の世帯数の比 $X_{t1} : X_{t2} : X_{t3}$ がその後の年次においても不変であると仮定していることである。第2に、誤差の評価が困難なことである。

これらの制約のうち、第1の制約に関しては、地域（調査区）への聞き取り調査を行って世帯の基本情報を得る方法、あるいは住民基本台帳を用いて無回答者の世帯構造と性別、年齢階級を把握する方法などによって対処することが可能であろう⁴⁾。こうした補足的な情報を活用することによって、世帯数の比率を調整していくことは困難ではないと思われる。第2の制約は現時点では理論的に未解決の問題であり、今後の研究の成果をまちたい。

さらに付言すれば、国民生活基礎調査の世帯数などの推計に本研究の補正方法を導入する場合、従前の方法による推計値と新しい方法による推計値をどのように接合するかという問題が生じる。この点についても研究すべき余地があるものと考えられる。

V 結 語

国民生活基礎調査による世帯総数、単独世帯数などの推計値を実態に近づけるといって、本研究で提示した補正方法は有効な方法であるといえる。しかし、この補正方法を実際に利用する場合、住民基本台帳などから無回答世帯についての補助的情報を入手してこれを活用する

ことが必要である。また、過去の推計値とどのように接合するかという課題を解決しなければならない。

謝辞

本研究の機会を与えてくださった(財)厚生統計協会、研究会のメンバーである早稲田大学政治経済学部西郷浩教授と厚生労働省国民生活基礎調査室の皆様へ感謝申し上げたい。

文 献

- 1) 新田功, 西郷浩, 他. 国民生活基礎調査における無回答データ等の影響を考慮した調査設計に関する研究報告. 国民生活基礎調査における無回答データ等の影響を考慮した調査設計に関する研究会, 2004.
- 2) 村山令二, 鈴木健二, 石井大, 他. よくわかる標本調査法—厚生統計で学ぶ標本設計の理論と実践. (財)厚生統計協会, 2004.
- 3) 杉本明子. 社会調査の基本. 朝倉書店, 1984.
- 4) 林文, 山岡和枝. 調査の実際—不完全なデータから何を讀みとるか. 朝倉書店, 2002.
- 5) 西郷浩. 無回答補完の手法と推定量の精度評価. 早稲田政治経済学雑誌 2002; 349: 155-81.
- 6) 船岡史雄. 統計調査の方法と精度—国民生活基礎調査に基づく分析. 信州大学経済学論集 1995; 34: 1-15.