

全国調査での層化2段抽出法における最適な地点数割り当て法 JGSS-2000の地点数割り当ての評価

稲葉 太一

(神戸大学発達科学部)

An Optimal Assignment of Number of Examination Points at Two-stage Stratified
Sampling by National Social Survey

- Estimation of assignment of number of examination points at JGSS-2000 -

Taichi INABA

By National Social Survey, we often use two-stage stratified sampling that divide whole country into 6 blocks(Hokkaido-Tohoku, Kanto, Tyubu, Kinki, Tyugoku-Sikoku, Kyusyu) further divide 3 population-size(13-big city, other city, suburban districts) . This paper reports difference of following three methods related assignment of number of examination points. First method was proposed by Central Research Services, INC, used JGSS-2000. Second method was proposed by Central Research Services, INC, used JGSS-2001. Third method is proposed by us. Really survey is carried out at 1.5 year later, we estimate really population used by Resident-Fundamental-Register and Population-Census data already published.

Key words: JGSS, two-stage stratified sampling, Assignment of number of examination points

全国調査において、全国を6つのブロック(北海道・東北、関東、中部、近畿、中国・四国、九州)に分け、その各々について更に3つの人口規模(13大都市、その他の市、郡部)に分類した層化2段抽出法は、比較的多く用いられている。この報告では、JGSSでの第2回予備調査以降でも行われている上記の層化2段抽出法における調査地点数の割り当てについて、3つの方法を検討する。1つ目は、中央調査社が従来から行っていた方法で、JGSS-2000に適用されている。2つ目は、中央調査社が新しく2001年度から用いている方法であり、JGSS-2001に適用されている。3つ目は、この報告で新しく提案する方法で、実際の調査は約1年半後であることから、既に得られている住民基本台帳人口と国勢調査人口のデータを用いて、実際の調査時点の人口割合を予測する方法である。

キーワード: JGSS、層化2段抽出法、地点数割り当て

1. はじめに

日本版総合的社会調査 (JGSS) は、1999 年に 2 回の予備調査を経て、現在、2000 年 10 月から 4 年間に渡る本調査が実施されている。

第 1 回予備調査を除けば、すべての調査は全国調査であり、いずれも層化 2 段抽出法を用いて 18 層 (6 地域 × 3 人口規模) の母集団人口を推定し、地点数を計算する方法を採用している。ここで、18 層とは、全国を 6 つのブロック (北海道・東北、関東、中部、近畿、中国・四国、九州) に分け、その各々について更に 3 つの人口規模 (13 大都市、その他の市、郡部) に分類したものである。

本報告は、上記の母集団人口を推定する方法として、中央調査社が 2001 年 3 月以前まで行っていた方法 (以下、中調従来法と呼ぶ) と、2001 年 4 月以降に行っている方法 (以下、中調新法と呼ぶ) を、実際の JGSS での地点割り当て数を基に検討する。

これらの検討においての本報告では、比較的安価で入手が容易であるという観点から、5 年に一度実施される国勢調査の結果と、毎年 3 月末時点で集計される住民基本台帳の結果の 2 つに限って議論する。得られるデータが増えればもっと良い方法があり得るだろう。また、調査方法決定時点で得られるデータと、実施日の間にどうしても約 1 年程度の乖離があるので、これを埋めるための方法 (以下、推定法と呼ぶ) についてもコメントする。

2. JGSS 調査概要

はじめに、JGSS の調査方法の内、今回の検討に関係する事項について説明する。JGSS の詳細については JGSS ホームページ[1] を見られたい。まず、第 1 回予備調査は首都圏と大阪府のみの調査であったので、今回の検討からは除く。また、第 2 回予備調査では、18 歳以上全体が調査対象であり、以後の JGSS-2000 (第 1 回本調査: 2000 年 10 月実施) と JGSS-2001 (第 2 回本調査: 2001 年 10 月実施) が 20 歳以上 89 歳以下 (90 歳未満) であることと併せて考慮すると、精度の面での比較検討が難しい。そこで、中調従来法の検討には JGSS-2000 を用い、中調新法の検討には JGSS-2001 を用いる事とする。

3. 母集団人口の推定について

3.1. 原資料

今回の報告での検討に関係する資料は以下の 5 つである。国勢調査については、平成 7 年度のデータは「日本の人口」[2] から、都道府県別市町村別の各歳人口を得た。ただし、得られたデータは総人口のデータである (日本人人口ではない)。平成 12 年度のデータは、総務庁統計局ホームページ[3] から、全国のデータを得た。これは、総人口も日本人人口もある。住民基本台帳については、「住民基本台帳人口要覧」[5][6][7] の平成 7, 12, 13 年度版から都道府県別市町村別の 5 歳きざみのデータを得た。

3.2. 日本人人口と総人口

ここで、国籍の問題について整理しておく。国勢調査は、実際に日本に3ヶ月以上在住している（あるいは、在住する予定の）人全体が対象であるので、国籍は日本に限定されていない。ところが、住民基本台帳では、「住民基本台帳法第39条」に「国籍が日本人の者に限る」ということが明記されている。実際の運用が各自治体によって多少流動的であることは有り得ても、原則として日本人に限られていると考えられる。

一般的に言って、現在行われる社会調査において、抽出台帳を住民基本台帳や選挙人名簿とすると、調査対象は日本人に限られる。そこで、本報告では、日本人人口の推定に話を限定する。

4. 中央調査社の従来法と新法の説明

中央調査社では、2001年4月以降の調査に使用している「母集団」の年齢別人口に関する計算方法が従来と変わった。

従来は、住民基本台帳から各基本調査区単位での住民総数を得て、これに利用可能な最新国勢調査から得られた各歳人口比率を乗じて、「母集団」の各歳人口を推計していた。例えば、JGSS-2000（2000年10月実施）に対しては、1999年3月末の住民基本台帳人口総数に、1995年10月の国勢調査における年齢分布比率を乗じて算出されている。年齢スライドは行っていない。

これに対して、新しい方法は住民基本台帳から、市町村単位の人口総数のみならず5歳きざみの年齢分布を採用する。ただし、80歳以上は、住民基本台帳における括りが一括であるため、最新国勢調査の人口比率を年齢スライドさせて用いる。ただし、一部の年齢ではスライドを行わない。詳細は、中央調査社の「推定母集団」[4]を見よ。例えば、JGSS-2001（2001年10月実施）に対しては、2000年3月末の住民基本台帳人口データのうち20歳から79歳まではそのままのデータを採用し、80歳代は80歳以上の総人口に、1995年10月の国勢調査比率を乗じて推定している。

これらの両者を比較すると、年齢分布の推定において、従来法が常に「1995年の国勢調査」の抽出データを用いているのに対して、新しい方法は「5歳きざみなら1999年の住民基本台帳」の全数データを使うということであるので、全数である上に新しいデータを用いる可能性がある分だけ、新しい方法は必ず改善になっていると考えられる。

5. 推定法

中央調査社の方法では、いずれの方法であってもJGSSの場合には実施より約1年半以上前のデータのみから推定することになる。そこで、人口の18層毎の1年間の変化を「変化率」と定義し、これを推定することで、より現実のデータに近づけることが出来ると考えた。

まず、最も手軽な以下のような方法を検討してみた。平成7年と平成12年の住民基本台帳データは基本的に5年で5歳ずれていることから、5歳階級ごとに「5年変化率」を容易に求める事ができる。これを18層毎に算出しておき、その5乗根で「1年変化率」を推定した。次に、平成12年3月末のデータから、例えば20歳から24歳までの平成13年3月末の人口の推定方法は、平成12年の15歳から19歳までの人口の0.2倍に「15歳から19歳の1年変化率」を乗じ、平成12年の20歳から24歳までの人口の0.8倍に「20歳から24歳の1年変化率」を乗じて、これらを加えることで推定した。

6. 推定精度の検討

3つの方法を比較検討する。ここで、実際のJGSSの実施は10月であるが、比較検討のために、すべて4月に実施されたと考える事にする。実際は、更に半年後に実施されているので、推定法に不利な状況での比較となっている。

6.1. 中調従来法の精度評価

中調従来法の精度を検討する。従来法で調査地点数を求めたJGSS-2000では、次の表1のような割り当てであった。ここで、JGSS-2000の真の母集団は、2000年4月の人口である。これは正にJGSS-2001の表2を作成するために用いられた「住民基本台帳平成12年度版」の母集団と全く同じである。このとき、80歳代については国勢調査の情報を用いて年齢分布から推定を行っている。実は、中央調査社が行っている推定方法は、各市町村ごとに住民基本台帳人口と国勢調査年齢分布を用いて詳細に行っており、簡単に修正できるものではない。従って、80歳代を簡単に取り除く事は出来ないので、80歳代のずれは無視することにする。

そこで、表1と表2を、各18層の人数を全調査人数で割って比率に換算し、18層毎に差の絶対値を取り和を計算したところ、0.862% (300地点で換算すると、2.59地点)となった。実際、JGSS-2000とJGSS-2001の間で、3地点の変更があった。(北海道・東北の13大都市、関東のその他の市、九州のその他の市の、3地点が1つ地点数を増やし、いずれも郡部が3地点減らしている。)

表1 ブロック・市郡規模別（18層）の母集団数・標本数・調査地点数（JGSS-2000）

	JGSS-2000	SIZE 1	SIZE 2	SIZE 3	* TOTAL
		13大都市	その他の市	町村	
1	[北海道・東北] 7道県(2)	2,102,960 90(6)	6,020,350 285(19)	3,739,209 180(12)	11,862,519 555(37)
2	[関東] 8都県(4)	10,700,400 495(33)	16,508,164 765(51)	3,740,719 180(12)	30,949,283 1,440(96)
3	[中部] 9県(1)	1,634,257 75(5)	11,290,158 525(35)	4,906,146 225(15)	17,830,561 825(55)
4	[近畿] 6府県(3)	4,195,429 195(13)	9,523,335 450(30)	2,110,274 105(7)	15,829,038 750(50)
5	[中国・四国] 9県(1)	834,684 45(3)	5,528,151 255(17)	2,824,111 135(9)	9,186,946 435(29)
6	[九州] 8県(2)	1,742,678 75(5)	5,407,678 255(17)	3,611,327 165(11)	10,761,683 495(33)
計	母集団数 標本数(地点数)	21,210,408 975(65)	54,277,836 2,535(169)	20,931,786 990(66)	96,420,030 4,500(300)

表2 ブロック・市郡規模別（18層）の母集団数・標本数・調査地点数（JGSS-2001）

	JGSS-2001	SIZE 1	SIZE 2	SIZE 3	* TOTAL
		13大都市	その他の市	町村	
1	[北海道・東北] 7道県(2)	2,199,106 105(7)	6,195,110 285(19)	3,823,384 165(11)	12,217,600 555(37)
2	[関東] 8都県(4)	10,999,069 495(33)	17,138,198 780(52)	3,805,192 165(11)	31,942,459 1,440(96)
3	[中部] 9県(1)	1,678,685 75(5)	11,619,542 525(35)	5,002,788 225(15)	18,301,015 825(55)
4	[近畿] 6府県(3)	4,310,465 195(13)	9,799,730 450(30)	2,166,563 90(6)	16,276,758 735(49)
5	[中国・四国] 9県(1)	863,645 45(3)	5,669,310 255(17)	2,875,526 135(9)	9,408,481 435(29)
6	[九州] 8県(2)	1,806,228 75(5)	5,878,028 270(18)	3,707,182 165(11)	11,391,438 510(34)
計	母集団数 標本数(地点数)	21,857,198 990(66)	56,299,918 2,565(171)	21,380,635 945(63)	99,537,751 4,500(300)

6.2. 中調新法の精度評価

JGSS-2001での割り当ては、2000年3月末の住民基本台帳データを元に計算されるが、本来の母集団は2001年4月のデータである。これらの違いは、まさに平成12年度住民基本台帳と平成13年の住民基本台帳のずれと言える。ただし、20歳から79歳までは実際の人口が分かるが、80歳代は分からない。そこで、80歳代の推定方法による誤差を除き、取り敢えず、20歳から79歳までのずれを見ることにした。

精度の比較方法については、中調従来法と同様に、2つのデータで共に18層の比率を求め、その差の絶対値を加えた所、0.297% (300地点で換算すると、0.89地点)であった。80歳代が考慮されていないとは言え、従来法の3分の1程度のずれになっていることから、JGSSの実際の地点数の計算に関しては改善になっていると思われる。

いずれにせよ、真の母集団とのずれが、300地点の割り当てに換算して1地点以下となっていることから、これ以上の精度を達成する事の価値がどちらかというと低下しているレベルに達していると考えられる。

6.3. 推定法の精度評価

最後に、推定法についても精度評価を行う。平成7,12年度の住民基本台帳人口から、平成13年4月の人口を予測し、予測値と実際の平成13年度住民基本台帳人口とのずれを評価した。住民基本台帳の実際の人口が分かる範囲で考えるため、20歳から79歳で比較した。18層での比率の差の絶対値を加えると、0.317% (300地点で換算すると、0.95地点)と中調新法よりも若干精度が劣るという結果となった。

ただし、5歳階級の中での人口分布の変動を考えると、このような荒い推定法でも1年後の予測を全くしていない中調新法とほぼ同レベルであったことは、むしろ前向きに評価できると考えている。

7. まとめ

たった一回ずつの実施例についての検討であるので、これをもって中調従来法と中調新法の絶対的な評価をしたとは言い難い。ただ、用いられているデータが、回収率が100%ではない国勢調査データから、ほぼ完全に捕捉されている住民基本台帳データになり、明らかに改善されている中調新法の性能が、地点割り当て数の決定に際しても有効に機能している事がほぼ確かめられたので、今回の中央調査社の変更は、前向きであったと思われる。

また、上記の中調新法と推定法では、推定法が若干劣るという結果であったが、今後、国勢調査の各歳人口のデータを利用することで、推定法の精度をより向上させる可能性があることと、実際のJGSSの実施は1年後ではなくて1年半後なので、推定法の利点はより大きいと考えている。具体的に、より良い推定法を構成することが今後の課題である。

[参考資料]

- [1] JGSS ホームページ, <http://www.jgss.daishodai.ac.jp/>
- [2] 日本の人口, 平成7年国勢調査最終報告書, 総務庁統計局編
- [3] 総務庁統計局ホームページ, 第3表「年齢(各歳), 男女別人口(日本人) 都道府県別市町村別」
<http://www.stat.go.jp/data/kokusei/2000/kihon1/00/hyodai.htm>
- [4] 総務庁統計局ホームページ, 第4表「年齢(各歳), 出生の月(4区分), 男女別人口(日本人) 全国」
<http://www.stat.go.jp/data/kokusei/2000/kihon1/00/hyodai.htm>
- [5] 住民基本台帳人口要覧, 平成7年度版, 自治省行政局編
- [6] 住民基本台帳人口要覧, 平成12年度版, 市町村自治研究会編
- [7] 住民基本台帳人口要覧, 平成13年度版, 市町村自治研究会編
- [8] 中央調査社の「推定母集団」-2001年度から年齢スライド方式で計算- 中央調査報
(No.522) <http://www.crs.or.jp/5221.htm>