

近年 5 回の国民性調査の標本設計と 標本精度について

統計数理研究所 前田忠彦・中村 隆

(受付 2000 年 1 月 11 日; 改訂 2000 年 3 月 21 日)

要 旨

日本人の国民性調査近年 5 回 (第 6 次～第 10 次) の全国調査について、標本設計を説明した上で、二つの観点から標本精度に関する基本的な資料を提示した。一つは標本の性・年齢構成であり、もう一つは標準誤差の大きさである。標本の性・年齢構成の検討では、計画標本と回収標本の両方について、日本人母集団の性・年齢構成との比較を行い、標本構成の偏りが結果数値に与える影響を評価するために、性・年齢で事後層別をした標本の重み付き集計と単純集計の差を吟味した。主たる知見は次の通りである。1) 回収標本では計画標本に比べて母集団の性・年齢構成からの乖離が大きくなり、近年 2 回の調査で特に大きい。2) 重み付き集計と単純集計の差は多くの場合に 1%未満であり、単純集計数値が受けた影響は概して小さいと言える。ただしその差は第 10 次調査で大きい。差が大きくなる項目は、複数の調査間で一致している。標準誤差の検討では、各年次の標本設計が層化確率比例復元 2 段無作為抽出であると仮定して母比率の推定量とその標準誤差を計算し、以下の知見を得た。3) 標本サイズが小さいので第 10 次調査では標準誤差が他の調査年次より大きめである。4) 単純無作為抽出の場合に対する標準誤差の倍率は、第 6 次と第 7 次では 1.1～1.2 倍、第 8 次と第 9 次では、1.2～1.3 倍である。第 10 次は後者のグループに近い。5) 比率推定量の単純集計との差は、回収率が低いほど大きくなる傾向がある。

キーワード：日本人の国民性調査、層化 2 段無作為抽出法、標本誤差、推定量の分散、標準誤差の倍率、バイアス。

1. はじめに

本稿では、近年 5 回の「日本人の国民性調査」における標本設計の概要を解説した上で、

- 1) 計画標本及び回収標本の性・年齢構成、特に母集団の性・年齢構成との比較
- 2) 各調査年次の標本設計を層化 2 段抽出法と仮定した場合の推定標準誤差

という 2 つの側面にわたり、同調査の標本精度に関係する基本的な資料を時系列的に提示し、今後の結果数値の利用に当たって考慮すべき基本的な情報を提供する^{*)}。以下では、研究の背景と、本稿の構成 (各章での具体的な検討事項) について述べる。

1.1 研究の背景

1953 年以来 5 年おきに実施され、1998 年秋に第 10 次の全国調査が行われた「日本人の国民

性調査」は、全国の20歳以上の有権者を対象とする標本調査である。調査の方式^{※2)}を各回で統一し(同一調査方式)かつ同一質問項目を繰り返し利用する継続調査^{※3)}を基本姿勢として、調査回数を重ねてきた。(以下、たとえば「第6次全国調査(1978)」のことを本文中では「第6次」のように書き、「全国調査」という語は省略する。調査年は初出時以外必ずしも示さない。)

同一調査方式による継続調査とは言っても、第9次(1993年)からは実査の方式に1つの大きな変更があった。第8次(1988年)までは、統計数理研究所が全国各地の大学に依頼し、大学生を調査員として(いわば自前で)調査を行っていたところを、第9次からは、実査(調査地点における選挙管理委員会での調査対象者の抽出と現地調査)の部分を外部の専門調査機関に委託して行うようになった。第10次(1998年)も、サンプリングの一部(調査地点の抽出及び一部の調査地点における調査対象者の抽出)を統計数理研究所が直接行った以外は、実査の部分を中心に外部の専門調査機関に委託した。実査を担当する機関が違えば、実際の訪問面接調査時の諸条件を含めた調査プロセスが異なる^{※4)}。

また、サンプリング方法について言えば、こうした全国調査を行う場合の標準的方法として、(全国の有権者を母集団とする)層化多段抽出法という大枠は各回で変わりなく用いられているものの、細部の手順などは異なっている。近年の比較的大きな変更点は、第8次以降、従来よりも層別を細かく行わなくなったという点が挙げられる。この点は2.4節で詳述する。

さらに、よく知られる通り近年社会調査における調査不能率が増大しており、この点は国民性調査も例外ではない。第6次以降の回収率を順に挙げると、第6次(1978年)73%、第7次(1983年)74%、第8次(1988年)61%、第9次(1993年)69%、第10次(1998年)64%である。第8次で従来に比べて10ポイント以上回収率が急落した。こうした回収率の低下が一過性のものとは考えられないこと、そうした中で回収率を保つための良質の調査員を自前で調達することが困難になっていることが、第9次以降に実査を専門調査機関に委託することになった直接の動機である(坂元(1995a))。

こうした調査方式(「実査担当機関」「標本設計における層別の方法」)の変更や「回収率」の低下は、調査結果に何らかの影響を与えている可能性がある。そこで本研究では、第6次(1978年)以降の近年5回の調査について、標本精度に関係する数値を提示し、調査年次間で適宜比較しながらその特徴をできる限り客観的に要約する^{※5)}。

第6次から第10次の調査でどのような条件の違いがあったかを表1にまとめた^{※6)}。5回の調査の間では、標本サイズが異なっており(計画標本サイズ、回収標本サイズとも)、これは当然標準誤差等の精度指標の大きさに影響を与える。したがって異なる調査間の比較を行う際には、標本サイズの影響をある程度抑制した指標を利用する(5.6節などを参照)。そのため表1では標本サイズを明示的な条件差としていない。

これら条件の違いのうち、回収率の低下の影響は4章に示す通り、第一に回収標本の性・年齢構成の偏りに顕著に現れている。両者は軌を一にする問題であるとも言える。性・年齢など

表1. 近年5回の国民性調査における各種条件の差。

調査年(略号)	条件1 実査担当機関	条件2 層別の方法	条件3 回収率
1978(第6次)	統計数理研究所	方式C ^{a)}	73%
1983(第7次)	統計数理研究所	方式C ^{a)}	74%
1988(第8次)	統計数理研究所	方式D	61%
1993(第9次)	専門調査機関A	方式D	69%
1998(第10次)	専門調査機関A	方式D	64%

注a) 第6次と第7次の層別の方法は完全には同じではないが、類似している。

の基本属性項目に関する偏りが、基本属性以外の調査項目の結果数値にどのような影響を及ぼしているかについても、検討材料を示しておくことが必要と考える。

さらに、調査結果の精度は具体的には集計結果（平均値、比率）の標準誤差の大きさにより評価される場合が多いから、各回の標本設計の下で、標準誤差等の大きさを具体的に評価しておくことも必要である。標本誤差の大きさを調査年次間で比較しておくことも望まれるが、異なる標本設計の下での調査の標本誤差の比較は、たとえ同一項目であってもやや難しい問題を抱えている（5.6 節に後述）。このため、大まかな目安を与えるに止める。

1.2 各章での検討事項

本稿は、以下のように構成される。

まず第 2 章では、第 6 次以降の全国調査の標本設計の概要を述べる。本稿の検討事項と関連する部分についてはやや詳しく紹介する。第 3 章では以下の分析に利用した調査項目を説明する。第 4 章は、第 1 の検討側面である「計画標本、回収標本の性・年齢構成」の経年変化、特に母集団の性・年齢構成からの乖離を述べ、関連して、その乖離が結果数値にどのような影響を与えたかを検討する。第 5 章では第 2 の検討側面である「標本誤差の評価」に関して本稿の検討方法を説明する。第 6 章は第 5 章で述べた方法による標本誤差の記述である。層化 2 段抽出を仮定した場合と単純無作為抽出を仮定した場合の母比率の推定値の差、及び標準誤差の記述とその年次比較に焦点を当てる。第 7 章では以上の検討事項を補足し、残された検討課題について述べる。また近年の国民性調査の標本構成と標本誤差について得られた知見をまとめる。

2. 「日本人の国民性」全国調査の標本設計

「日本人の国民性調査」(全国調査)の標本抽出は、全国の有権者を母集団とする層化多段抽出法を用いて行われてきた。ここでは第 6 次以降の 5 回の調査の標本設計について共通の事項をまず述べ、続いて第 6 次・第 7 次と第 8 次以降で異なる層別の方法について（第 8 次以降の層別方法をやや詳しく）述べる。具体例を出す場合には第 10 次を中心とする近年の調査によるが、他の調査年次でも本質的な相違はないものと考えて良い。なお、第 10 次の標本設計の概略は中村 他 (1999) に記されている。

本章の記述は国民性調査の標本設計に関して従来発表されてきた資料に依拠しているが、説明内容は、1995 年 SSM 調査 (社会階層と社会移動に関する全国調査^{*)}) の標本設計を解説した中村 (1998) と一部重複している。

2.1 調査票の種類と計画標本のサイズ

国民性調査では、1973 年の第 5 次から、調査票を 2 種類用意し、標本を二分して (第 5 次のみ 2 対 1 の割合、以降は 1 対 1 の割合で分割) 実施する方式をとっている。すなわち、従来からの継続質問項目を中心とする「K 型 (過去型) 調査票」、調査初期以降に現れた新しい意識動向に対応するべく新しい質問を取り入れる (それらの質問についてはできる限り継続利用する) という方針の「M 型 (未来型) 調査票」の 2 種類の調査票を用いている。本稿で検討する第 6 次以降の 5 回の調査でも、K, M, 2 種類の調査票が用いられた。各調査票は調査地点内の対象者に対し一人おきに (対象者名簿の順序で交互に) 割り当てた。例えば第 10 次では、K 型は調査点の偶数番の対象者に、M 型は奇数番の対象者に用いた。

計画標本サイズは第 6 次から第 10 次まで順に、5400, 6000, 6000, 5400, 4200 である。

2.2 母集団

調査対象集団は、調査年前年の12月31日現在で満20歳以上の日本人である。これに対する母集団は、基本的に有権者名簿に基づく全国の有権者となっている（ただし、島嶼部など一部の遠隔地居住者はあらかじめ除かれている）。標本抽出で用いた母集団の大きさとしての有権者数は、基本的に調査前年9月1日現在の数字（一部の町村では前々年9月1日現在の数字）である^{#8)}。厳密には調査年での精確な母集団の大きさではない。

2.3 標本設計方法の概略

国民性調査では、他の全国規模の調査の多くと同様に、自動加重標本に近づくように計画標本段階で下に述べるような方法をとっている^{#9)}。

各層に標本を比例割当し、層内では調査地点を（その大きさに比例する確率で）確率比例復元抽出して、各地点の標本サイズを同一にすれば、得られる標本は自動加重となる（5.4節参照）。ただし、割当数の丸めや調査地点ごとの回収率の違いにより、厳密な自動加重標本を最終的に得るのは実際上困難である。困難ではあるが、なるべく自動加重標本になるよう計画しておけば、簡便法として回収標本に関して単純集計が許容されると期待できる。

標本設計手順の概略は、具体的には以下の通りであり、これは各回に共通な枠組である。

- [1] 市区町村別人口（有権者数）等の情報を収集する。
- [2] 市区町村に対し層別を行う。層の数を L 、各層を h と表記する。
- [3] 総標本サイズ n と総地点数 m を決定する：第6次～第10次とも $m = 300$ 。
- [4] 各層の標本サイズ n_h の決定： n を層の大きさにしたがって各層に比例割当する。
- [5] 各層の地点数 m_h の決定： m を層の大きさにほぼ比例するように各層に割り当てる。
- [6] 地点毎の標本サイズの決定：地点毎の標本サイズ n_{jh} は、まず n_h/m_h の整数部分の数を最低標本サイズとして割り当て、残余の分は重複なくランダムにいくつかの地点に追加割り当てする。

[4][5]のように標本サイズと地点数を共に層へ比例割り当てすることで、全層を通じて各地点の標本サイズがほぼ一定になる。ただし[6]のように若干の端数処理が必要となる。調査地点の抽出、最終抽出単位の抽出の手順についてはそれぞれ、2.5節と2.6節に別記する。

2.4 層別の方法

2.4.1 第6次・第7次調査の方法

第6次、第7次の層別は、全国の市区町村を、区・市・郡の別、地方性、人口規模を考えて層別する方式である。第6次では郡部の一部について、産業構成（第一次産業従事者人口）も考慮している。第6次の層別の詳細については統計数理研究所国民性調査委員会（1982），pp. 424-427，第7次の層別の詳細については統計数理研究所国民性調査委員会（1992），pp. 474-477に記されている。

具体的には、全国の市区町村が第6次では計68層、第7次では計56層に区分された^{#10)}。このように層別を細かくすると、層のサイズが小さくなるために、地点あたりの標本サイズが層間で変動し過ぎることが起こる^{#11)}。この点が実査管理の上からは問題になり、このことが第8次以降で層別の方法を変更した理由である。

2.4.2 第8次～第10次調査の方法

第8次から層別の方法を変更し、第9次、第10次も第8次での変更を踏襲している。この層別の方法の狙いは統計数理研究所国民性調査委員会（1992），p. 483に書かれている。ここでは

表2. 第10次全国調査の標本設計の概要。

層の名称	母集団人口 (有権者数)	地点総数 (市町村数)	割当		計画標本サイズ	
			地点数	K型	M型	合計
全国計	98,019,819	3,287	300	2,075	2,125	4,200
区 部	21,588,792	150	68	449	476	925
市部(人口20万人以上)	24,822,046	92	75	525	539	1,064
(人口20万人未満)	30,071,061	555	92	644	645	1,289
郡 部	20,747,838	2,462	63	441	448	889
沖 縄	790,082	28	2	16	17	33

その概略を説明する。

層別として、全国の市区町村(に登録されている有権者)を大きく5つの層に分けた。すなわち、「区部(政令指定都市の区)」「人口20万人以上の市部」「人口20万人未満の市部」「郡部」「沖縄」である。市区町村の属性としての人口規模は(有権者数ではなく)、直近の1985年、1990年、1995年の国勢調査人口である。

各層内では市区町村をある順に並べる。これは、層の数が少なく層別が一見大まかであることを補うために調査地点(以下、単に地点と呼ぶことがある)を系統抽出し、層別を細かくしたのと同様の効果を期待してのことである。区部と郡部は地方公共団体コード順、その他の層ではまず各層内で市町村を都道府県別にグループ化し、さらに各都道府県内では市町村を人口規模の昇順または降順に並べる(都道府県が変わるごとに昇順、降順、昇順、...を繰り返す)。このようにして性質が類似した市区町村の並んだリストから、2.5節の系統抽出を行うことで、並べ替えが一種の層別の効果を持つ。

表2に、上記の層別に関する第10次の標本設計の概要を例示する。

2.5 調査地点の抽出

ここでは第8次以降の調査地点の抽出方法について述べる。

これまで調査地点ということばを慣用的に使ってきたが、これはどちらかといえば実査管理上の概念であり、1調査員が担当する範囲という程度の意味である。標本設計上は抽出単位を明確にする必要がある。

この方式の場合、第1次抽出単位(PSU, Primary Sampling Unit)が市区町村、第2次抽出単位が投票区、第3次抽出単位が調査単位集団、最終抽出単位が個人と考えられる。ここで、調査単位集団とは「地点の標本の大きさ×抽出間隔」の大きさの集団を想定したものである。この概念は最終抽出単位の抽出に系統(等間隔)抽出法を利用していることから生じる。

上のように述べたが、実際には諸抽出単位を明確に区別しながら抽出したわけではなく、次のような系統抽出を利用して調査地点を決めた。

- [1] 層別の節(2.4.2節)で述べたように各層内で市区町村を並べる。各市区町村内では投票区が何らかの順にすでに並んでいて、さらに投票区内では有権者が名簿に記載されている順に並んでいることになる。
- [2] 層のサイズ N_h を層の割当地点数 m_h で割った値 d_h をその層の調査地点のための抽出間隔とする。0以上 d_h 未満の一樣乱数を発生させ、各層についての開始値 c_h として定める。(d_h , c_h は整数に丸めない。)
- [3] 層の開始値 c_h および c_h に割当地点数に達するまで順次抽出間隔を足していった値 $c_h + d_h$, $c_h + 2d_h$, ..., $c_h + (m_h - 1)d_h$ をそれぞれ整数に切り上げ^{#12)}、この順番に当たる有

権者を各調査地点の調査開始者と考える。この開始者を含む投票区が第2次抽出単位、その投票区を含む市区町村が第1次抽出単位となる。また、調査地点のいわゆるスタート番号——当該投票区の有権者名簿の先頭から何番目から抽出をはじめるか——が自動的に決まることになる。

2.6 調査対象者の抽出

調査対象者（最終抽出単位）の抽出は、抽出された市区町村の選挙管理委員会に出向き、上のように決定した投票区のスタート番号から、指定された抽出間隔おきに有権者個人を抽出した（等間隔抽出法）。実査上調査地点の地域的広がりをあまり大きくしないことと、有権者名簿上の並びでの性別の自己相関を考慮して、第8次以降の抽出間隔は、区部及び人口20万以上の市部（沖縄の那覇市を含む）では21、それ以外では11と奇数とした。

3. 検討に使用した項目とカテゴリ

第4章、第6章では、標本精度に関係する指標を項目（カテゴリ）ごとに算出し結果を吟味する。検討する項目は、原則として「利用可能な」ほとんど全ての項目とカテゴリであり、具体的には以下の方針で選定した。

1) 個々の調査年次の中で検討する場合：

- ・各調査のK型・M型調査票とも、自由回答法による一部項目を除く全ての質問項目、及び属性項目のうち性、年齢、学歴、職業を使用する。
 - 属性項目と択一式項目の選択肢のうち、“その他”及び“D.K.”を除く全選択肢を0-1型（該当すれば1、しなければ0）の変数と見て集計対象とし、以下「カテゴリ」と呼ぶ。
 - 複数回答の項目では各選択肢ごとに、「該当」（各選択肢を選択する）のみを一つのカテゴリと見なす。やはり“その他”および“D.K.”は除く。
- 以上について、用語としては「項目」の中にいくつかの「カテゴリ」（集計のための0-1型の変数）がある、という形の使い分けをする。
- ・“その他”“D.K.”を含まない項目からは、カテゴリ番号の大きいカテゴリを一つ除外する。これは、項目内で比率の数値が従属な（和が1になる）カテゴリが含まれないようにするための配慮である。
- ・単純集計の結果が2%未満あるいは98%より大のカテゴリは、外れ値的に働く場合が多いことを考慮して、分析対象から除外する。

2) 調査間で共通の項目によって検討する場合：

K型、M型の区別を取り払い、1)による全カテゴリの内、属性項目を除き、複数の調査年次に共通のカテゴリを選定する。その際、各調査内のK型、M型双方の調査票で使用されている場合は、標本サイズの条件をできる限り斉一に保つため、K型とM型を合計した標本ではなくK型の標本の結果を利用する。

表3. 各調査年次及び調査票毎の分析対象項目・カテゴリ数。

調査年次 調査票	第6次		第7次		第8次		第9次		第10次	
	K型	M型	K型	M型	K型	M型	K型	M型	K型	M型
項目数	43	48	47	49	46	48	51	63	50	62
カテゴリ数	132	145	163	194	164	168	172	219	179	211

以上1)により分析対象となった項目、カテゴリ数を、一覧にして表3に示す。また2)により、全5回の調査年次に完全に共通なものは、28項目84カテゴリとなった。多くはK型調査票に含まれる項目である。

4. 性・年齢構成に関する標本と母集団の比較

4.1 検討目的

本章では各回の標本の性・年齢構成に着目し、母集団人口（調査年10月1日現在の20歳以上の日本人の推計人口）の性・年齢構成との比較を行う。これは調査の精度を、まず標本が「性・年齢という最も基本的な属性に関して母集団を適切に反映しているか」という観点から検討するためである。すなわち、この意味での精度が近年の5回の調査でどのように推移したかを検討し、国民性調査の計画標本と回収標本の性格に関する一側面を記述する。

性・年齢については標本抽出の時点で判明するため、計画標本と回収標本の両方について上記を検討する。更に母集団の性・年齢構成からの乖離が集計結果にどのような影響を与えるかを検討するために、標本を性・年齢について事後層別し、母集団における性・年齢の層のサイズで重み付けた集計結果と単純集計結果の差を検討する。

なお本研究では年齢区分を、20～24歳、25～29歳、..., 65歳～69歳、70歳以上の5歳刻み11段階としており、これは国民性調査の調査項目#1.2による。

4.2 各次調査における標本の性・年齢構成

図1は計画標本の性・年齢別の構成比率を男女別及び男女計（左列）について、調査年次毎に示したものである。図中の実線は、母集団としての（各調査年の10月1日時点の）推計日本人人口（総務庁統計局の人口推計資料による）における性・年齢構成を示す。また図2は回収標本についての同様の図示である。

図中の破線は単純集計（単純無作為抽出SIを仮定した集計；後掲(5.8)式を0-1データに適用した場合； $\hat{p}_{SI}$ ）、点線は層化確率比例復元2段階抽出STPPR-SIを仮定した重み付き集計（(5.6)式を0-1データに適用した場合； $\hat{p}_{STPPR-SI}$ ）を示す。計画標本では、ほぼ自動加重標本となっていることから点線と破線は殆ど一致し、図の上でも（解像度を超える差は出ずに）重なって表現されている。図2の回収標本では、地点ごとの回収率の違いにより自動加重の前提が崩れることから点線と破線は必ずしも一致しないが、ほぼ平行して描かれていると言ってよい。さらに点線のプロットの下上には、STPPR-SIを仮定して計算した標準誤差 $se(\hat{p}_{STPPR-SI})$ の1.96倍により、95%信頼区間を誤差棒の形で表示した。

図で視覚的に確認される内容を数値化しておくために、母集団としての推計日本人人口を真

表4. 推計日本人人口の性・年齢構成比率を真の分布とするK-L情報量。

	第6次	第7次	第8次	第9次	第10次
計画標本					
全体	0.0024	0.0015	0.0027	0.0137	0.0037
男・女	0.0030	0.0030	0.0037	0.0118	0.0053
回収標本					
全体	0.0013	0.0035	0.0027	0.0156	0.0185
男・女	0.0087	0.0096	0.0094	0.0205	0.0253

注) 全体の行は、男女を合わせた標本の11の年齢区分に関する計算、男・女の行は、性×年齢の22区分に関する計算である。

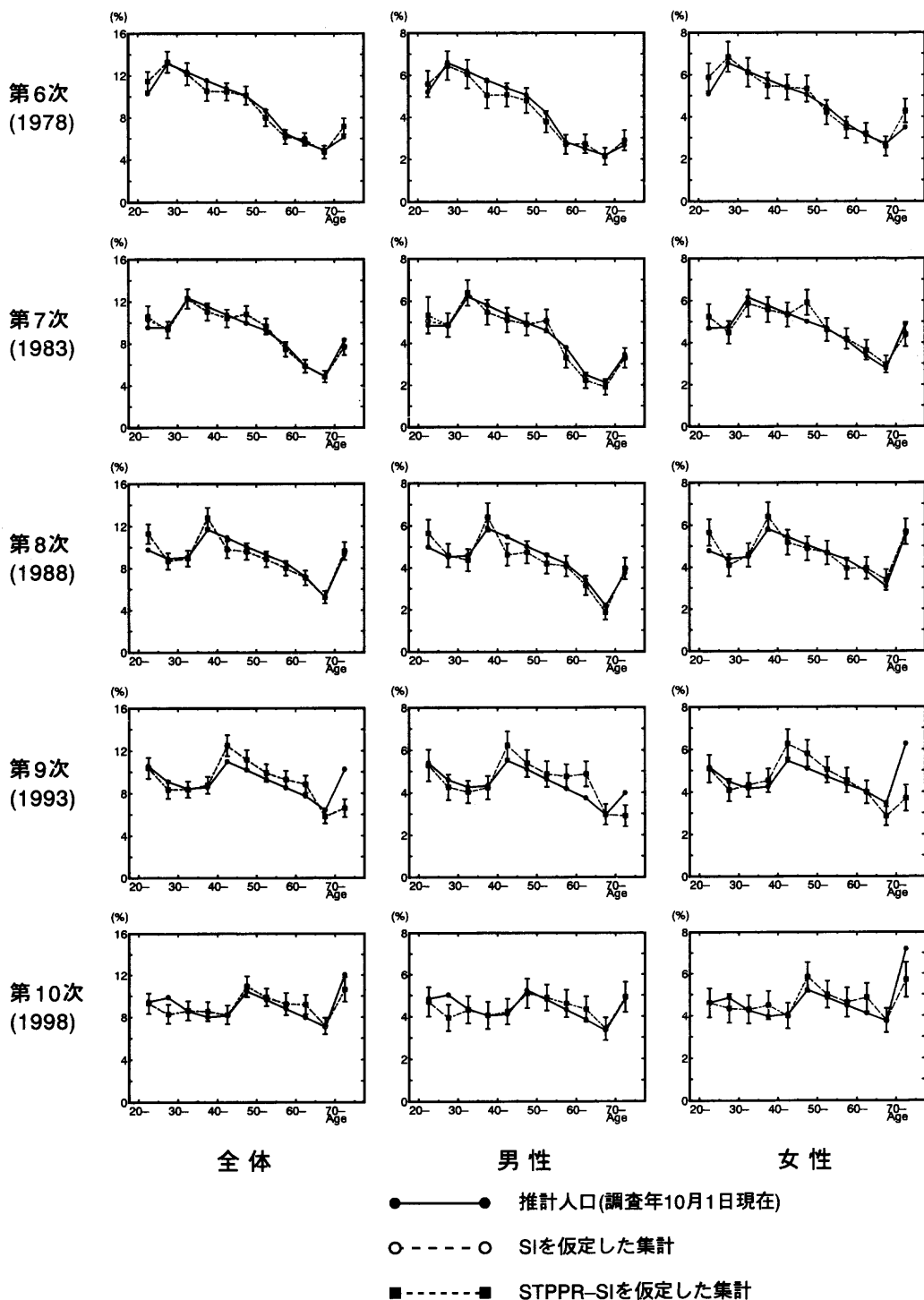


図1. 計画標本の性別・年齢構成の年次推移。

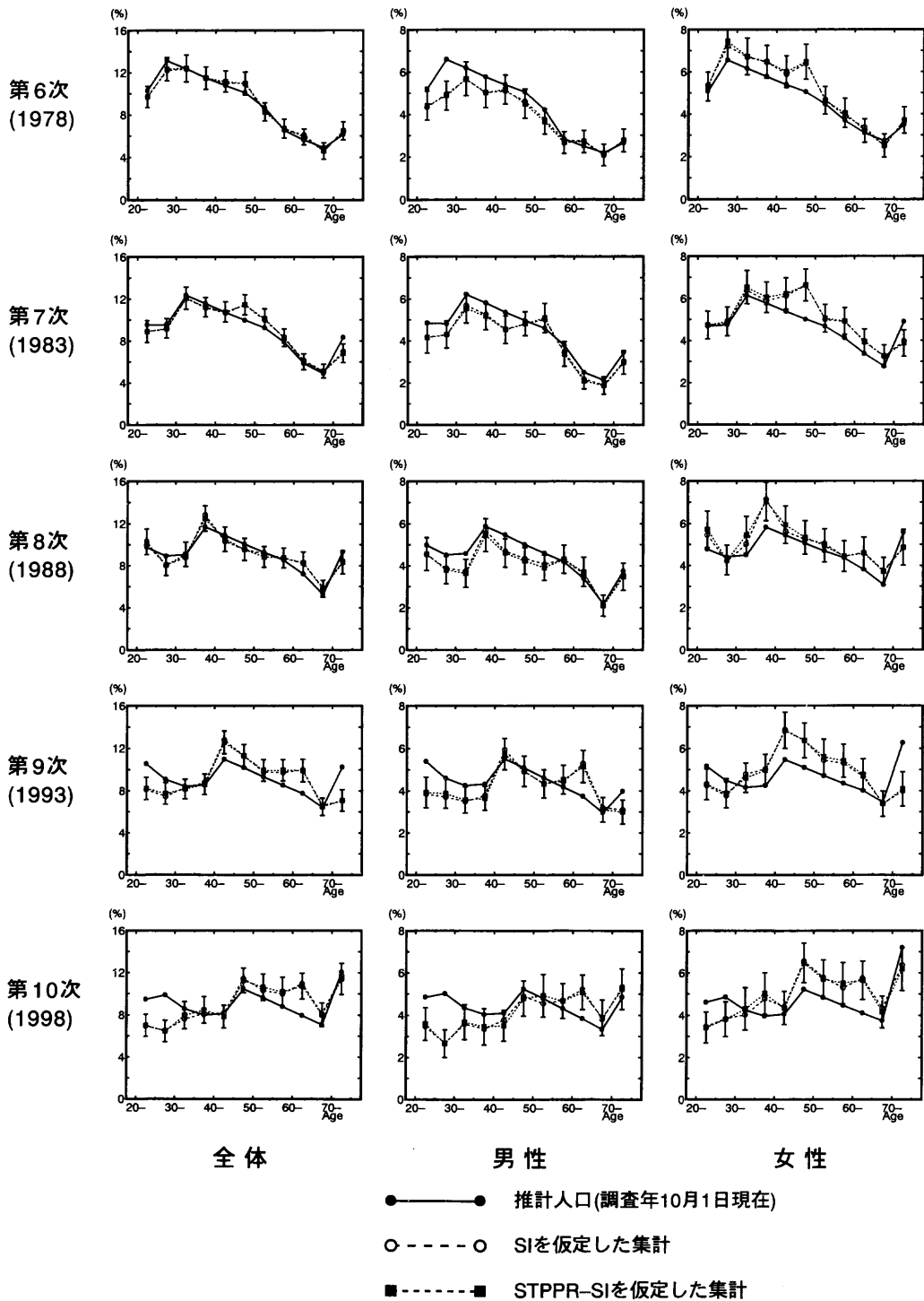


図2. 回収標本の性別・年齢構成の年次推移。

の分布とし、モデルと真の分布との近さを計る指標である Kullback-Leibler 情報量 (以下 K-L 情報量) を求めたものが表 4 である。

4.2.1 計画標本の性・年齢構成

まず、計画標本について次の特徴を指摘することができよう。

- (1) 日本社会全体の年齢別人口構成の変化に合わせて、計画標本でも高年齢層方向に少しずつ分布の重心がシフトしている。即ち、以前の調査よりも近年の調査のほうが、(計画標本の段階から) 全体の集計結果には高年齢層の属性・回答が相対的に強く反映される。
- (2) 第 9 次で母集団の性・年齢構成からの乖離がやや大きい他は、計画標本は母集団の性・年齢構成をよく捉えている。

第 1 点の指摘はもちろん国民性調査に限った傾向ではない。第 2 点の母集団の性・年齢構成との差については、各年次とも信頼区間から外れている性・年齢層が一部見られるものの、95% 信頼区間内に母集団の値が含まれる場合も多い。この乖離は次に述べる回収標本における性・年齢の偏りに比べれば小さい問題と考えられる。

4.2.2 回収標本の性・年齢構成

高年齢層の構成比率が近年の調査で相対的に大きくなる傾向は、回収標本で更に強調される。図 2 からは次のように指摘することができる。

- (3) 回収標本では各年次とも、(在宅率を反映して) 母集団より女性の構成比率が大きく、男性が小さくなっている。このうち第 6 次から第 8 次では年齢層を問わず女性が多く男性が少ないために、男女を合わせた全体では母集団の年齢構成を比較的よく反映したものになっている。
- (4) しかし、第 9 次、第 10 次では男女とも (しかし特に男性の) 若年層が母集団の構成比率より少なく、その分中高年層 (40 歳代以降) の女性と高年層 (60 歳代以降) の男性が多くなっている。その結果として全体の標本でも、母集団の年齢構成に比べて若年層が少なく、高年層が多く含まれている。

すなわち、回収標本の性・年齢構成は、母集団の性・年齢構成からは幾分ずれた標本になっており、その乖離は性・年齢に関してランダムに生じているのではなく、明らかな傾向性を有したものである、という点をまずは押さえておく必要がある。第 9 次と第 10 次の 2 回の調査で母集団の性・年齢構成からの乖離が顕著になることは、表 4 下半分の回収標本についての K-L 情報量にも示されている。第 6 次から第 8 次まではほぼ同水準であるが、第 9 次、第 10 次では K-L 情報量が異質に大きく、かつ第 10 次のほうが第 9 次よりも大きい。

計画標本の段階で自動加重標本となるような標本設計を採用していることから、国民性調査の結果としては従来から単純集計の数値が発表されている。上述の理由で、この数値は各調査年次共通に男性よりも女性の意見・属性を強く反映していること、第 9 次・第 10 次は中高年層の (特に女性のそれ) の意見をやや強く反映する数値になっていること、が推測される。近年 2 回の調査で中高年層の重みが大きくなることは、先に指摘した母集団の年齢別人口構成の変化に伴って必然的に生じる部分に加えて、若年層の調査不能の増加によってもたらされる回収標本における若年層の母集団構成からの下方のズレの大きさによって強化されている。

このような回収標本における母集団の性・年齢構成からの乖離の、調査結果の数値に対する具体的な影響を次節で簡単に検討する。

4.3 母集団の性・年齢構成情報の利用

4.3.1 単純集計のバイアス1の導入

前節で述べたように、(特に第9次・第10次において) 標本の性・年齢別構成が母集団分布から乖離していることは明白である。この乖離の結果、各項目の単純集計の数値は、真値より中高年層の意見を強く反映する方向に偏った結果になっていると考えるべきである。年齢間で差が全くない項目であればこのような偏りは生じないが、年齢差が大きい項目ではこの偏りが無視できない大きさになると予想しうる。

性・年齢構成の母集団分布からの乖離が、他の調査項目の集計結果に与えた影響を考察するために、一つの材料として次のような母集団の性・年齢構成情報を利用した重み付き集計値を求め、単純集計結果と比較することを考える。すなわち、2章で述べた層化2段抽出という枠を離れて標本を性・年齢で事後層別し、層毎の単純集計値を母集団の性・年齢構成比率で重み付けした加重集計を行い、この事後層別 (Post-STratification) による重み付き集計を $\bar{p}_{\text{PST}}$ と表記する。集計結果として公表されているのは単純集計 $\bar{p}_{\text{SI}}$ である。本稿では両者の差 $\bar{p}_{\text{SI}} - \bar{p}_{\text{PST}}$ を「(単純集計の) バイアス1」と呼び、必要に応じ Bias1 と略記する^{注13)}。当該回答カテゴリの母比率 (真値) が未知である以上、バイアス1が上で述べたような単純集計数値の偏りを適切に反映している保証は必ずしも得られないが、偏りの一つの目安と見なし得る数値ではある。例えば同一項目のバイアス1を調査年次間で比較すれば、第9次・第10次で乖離が大きくなった影響が各項目にどのように反映されたかの見通しを得ることができる。バイアス1の性質については7.1.1節で再度言及することとし、以下の節では標本の性・年齢別構成の母集団分布からの乖離が各調査項目に与える影響の指標として、バイアス1の分布を検討する。

4.3.2 バイアス1の絶対値の分布

表5は、バイアス1の絶対値の分布を数値要約したものである。この表では、年次別・各調査票別に、表3の項目 (ただしそこから性・年齢を除く) について計算した。

表5. バイアス1の絶対値とその対標準誤差比の分布に関する数値要約^{a)}。

	第6次		第7次		第8次		第9次		第10次	
	K型	M型	K型	M型	K型	M型	K型	M型	K型	M型
標本サイズ	2032	1913	2256	2173	1858	1824	1833	1905	1339	1341
カテゴリ数	121	134	153	183	153	157	162	208	169	200
絶対値の分布										
平均値	0.283	0.326	0.270	0.203	0.230	0.266	0.388	0.403	0.768	0.608
標準偏差	0.269	0.380	0.311	0.180	0.316	0.271	0.385	0.367	0.736	0.636
最大値	1.901	2.602	2.313	0.836	2.102	1.627	1.969	1.993	4.471	4.436
90%ile	0.654	0.643	0.539	0.460	0.454	0.598	0.789	0.936	1.878	1.393
第3四分位	0.389	0.397	0.380	0.298	0.275	0.350	0.541	0.549	1.104	0.838
中央値	0.191	0.245	0.202	0.153	0.147	0.190	0.259	0.310	0.506	0.423
第1四分位	0.115	0.104	0.075	0.070	0.064	0.090	0.104	0.110	0.264	0.191
10%ile	0.046	0.051	0.023	0.022	0.031	0.032	0.034	0.060	0.127	0.080
最小値	0.001	0.004	0.000	0.002	0.005	0.001	0.001	0.001	0.005	0.001
対標準誤差比										
平均値	0.314	0.356	0.315	0.242	0.242	0.280	0.421	0.430	0.680	0.545
標準偏差	0.277	0.371	0.317	0.193	0.299	0.259	0.422	0.361	0.623	0.558
1を超える割合(%)	2.5	3.7	2.0	0.0	3.3	3.2	9.3	7.2	20.8	14.5

注a) 各調査年次のK型、M型の有効カテゴリに基づいて計算した結果。絶対値の分布欄の単位は%。

まずバイアス1の絶対値(以下 $Abs(Bias1)$)の大きさそのものについて検討する。絶対値の分布について、表5の第9次までの結果で90パーセンタイルの値を見ると、いずれも1%未満となっている。つまり多くの項目では、単純集計の結果と性・年齢で事後層別した重み付き集計の結果には1%ほども差が出ない場合が多い、ということである。第10次の結果はやや異質で、パーセンタイルの値も平均値も他の調査に比べて上方にシフトした形になっている。

ここで絶対値の非常に大きいカテゴリに注目する。 $Abs(Bias1)$ は、第6次から第9次では大きい場合に2%前後の値となり、第10次では4.4%にも及ぶ。第10次で例示すれば、この最大値を与えるのは、項目#1.2b“婚姻・子”のカテゴリ‘既婚’であり、次が同じ項目のカテゴリ‘未婚’である。単純集計 $\hat{p}_{st}$ は前者を過大推定し($Bias1 > 0$)、後者を過小推定する($Bias1 < 0$)。この項目は対象者のライフ・ステージが大きく影響しているから、こうしたバイアスは当然の結果である。属性項目には他にも#1.3“学歴”(中学卒までのカテゴリを過大推定)や、#1.4c“職業”(主婦を過大推定、勤め人を過小推定)などのように、概して $Abs(Bias1)$ の大きくなるものが多い。つまり対象者の属性項目については、常に過大あるいは過小推定の可能性があると見ておくほうがよい。

むしろ注意すべきは、属性項目以外の項目で $Abs(Bias1)$ が大きくなるケースである。要約すれば、属性項目以外で $Abs(Bias1)$ が大きくなる項目も、その内容面での性格としては基本的に年齢差の大きい項目であると言える。例えば第10次K型の調査項目のうち、属性項目を除いて $Abs(Bias1)$ が最大のものは、#4.11“先祖を尊ぶか”に対する‘尊ぶ方’という回答カテゴリの3.301である。“先祖を尊ぶ”のは高年齢層に多い回答であり、単純集計は高年齢層の意見を本来より強く反映した集計値になっている。これと呼応して同じ項目の残りのカテゴリ‘普通’と‘尊ばない方’が順に $Bias1 = -1.693, -1.530$ となっている。

4.3.3 バイアス1の年次比較

次に、バイアス1の年次比較を考える。表5を見ると $Abs(Bias1)$ は第6次から第8次まではほぼ同水準で、第10次において突出して大きくなっており、第9次も第6次～第8次に比べて大きめである。

$Abs(Bias1)$ が第10次で大きい一つの理由として、第10次の標本サイズが他の4回の調査に比べて小さいために、全般に標準誤差が大きくなって(精度が落ちて)いることが挙げられる。しかし、そのことだけで説明のつく差ではないことは、表5下側「対標準誤差比」の3行に示された結果に明らかである。この対標準誤差比とは、年次間で標本サイズが異なることをある程度調整するために、 $Abs(Bias1)$ を推定標準誤差で割ったもので、ここでの標準誤差は便宜的に単純無作為抽出を仮定した場合のもの(5.13)式の平方根を利用している²¹⁴⁾。対標準誤差比の平均値を見ても、第10次の結果だけ他の4回の調査に比べて、異質に大きい。この指標が1を超える割合も、他の4次の調査では高々10%程度であるのに、第10次では表5のK型では20%になっている。詳細は示さないが以上の傾向は、検討項目を各年次に共通の84カテゴリに限定しても大まかにいって同様である。

この第10次の $Abs(Bias1)$ の大きさは、やはり回収標本の性・年齢構成の偏りを反映した結果であると考えるのが自然である。表5の第10次K型調査票で対標準誤差比が1を超える割合が高いのも、次のように解釈すれば説明がつく。「K型調査票に古くからの継続質問が多く、これは『伝統的対近代的』のような対立する価値観を表す選択肢の設定になった項目が多い。したがってM型調査票に比べて相対的に年齢差を生じやすいカテゴリが多いため、性・年齢構成の偏りを反映しやすかった」という解釈である。

以上のバイアス1に関する年次比較の結果を、次のように総括することができる。まず第8次までは、標本の性・年齢構成に関する偏りが結果数値に与える影響は、概して小さくなかつ

たと想定できる。図2で見たように、これら3回の調査では男女を併合すれば回収標本も母集団の年齢構成をほぼよく反映していたためであろう。一方第9次でこのバイアスがやや大きくなり、第10次では計画(回収)標本サイズの小ささだけでは説明のつかない大きなバイアス1の値が得られた。このことは、回収率の差に起因する性・年齢構成の偏りが、非標本誤差として近年の調査結果に対する影響を強めていることを示唆している。

5. 標本誤差の比較方法

本章では、標本誤差の計算及び調査間での比較に関する本研究での考え方を示す。誤差の評価式については標本調査法に関する標準的なテキスト(例えば Cochran (1977), 鈴木・高橋 (1998) など)の記述するところではあるが、本研究での標本誤差評価の考え方を説明する上で最低限の内容を述べておく。

一般に標本抽出の理論と実際の間には隔たりがあり、様々な問題が存在する(鈴木 (1996))。2章で述べた標本抽出方法も、地点の抽出と調査対象者の抽出に系統抽出を用いており、この標本設計に基づく標本誤差の評価は困難である。しかし、実際の標本抽出を無視して単純無作為抽出を仮定するにはやはり無理がある。したがって、単純無作為抽出でないどの程度の標本誤差になるかを見ておくためにも、計算が比較的容易で、かつ前記手順が近似的に実現していると見なし得るような標本設計を仮定して標本誤差を評価してみるのが妥当な線であると考ええる。以上の考慮に基づき、5章では、近年5回の国民性調査の標本設計が「層化確率比例復元2段無作為抽出法(STPPR-SIと略記)」であることを仮定して標本誤差の評価式を与える。

5.1 標本誤差推定の仮定

標本誤差推定の前提としたいくつかの仮定を列挙する。

- [1] 層別の方法は2.4節に既述の通りである。
- [2] 第1次抽出単位は投票区を想定し、第2次(最終)抽出単位は個人とする。
- [3] 各層からの第1次抽出単位(PSU)の抽出は、PSUの大きさに比例する確率比例復元抽出(PPR, Probability Proportional to size with Replacement)によるとする。
- [4] PSUからの最終抽出単位の抽出は、(非復元)単純無作為抽出(SI, Simple random sampling without replacement)によるとする。
- [5] 回収の有効・無効は、いかなる個人の属性・態度とも独立であると仮定する。

ここで、[3]については、実際は系統抽出によるので確率比例非復元抽出に近いともいえるが、非復元とすると計算が困難なこともあり、復元の確率比例抽出を仮定した。[4]についても実際は系統抽出である^{#15)}。[5]は強い仮定であるが、回収不能のことは無視し、回収された標本がもともと計画標本であるかのように考えるという意味である^{#16)}。

本稿では2段抽出法に関して、第1次抽出単位の抽出方法を示す略号(PPR)と第2次抽出単位の抽出方法を示す略号(SI)をハイフンでつなぎ、(各層ごとの)確率比例復元2段無作為抽出法をPPR-SIと略記する。更に層別を含めて考える際には、接頭辞としてST(Stratifed)をつけてSTPPR-SIとなる。

5.2 層別抽出

母集団の大きさを N とする。母集団が L 個の層に分かれているとし、第 h 層の大きさを N_h ($h = 1, \dots, L$) とすると、 $N = \sum_{h=1}^L N_h$ で、各層のウェイトは $W_h = N_h/N$ となる。第 h 層の第 i 要素の特性値を y_{ih} とすると、層の平均は $\mu_h = (1/N_h) \sum_{i=1}^{N_h} y_{ih}$ となる。層の平均 μ_h と母平均 μ の

間には $\mu = \sum_{h=1}^L W_h \mu_h$ という関係がある。ただし、層の区別を取り払って母集団の第 j 要素の特性値を $y_j (j = 1, \dots, N)$ と表記すれば、母平均は $\mu = (1/N) \sum_{j=1}^N y_j$ である。

以下の性質は層別抽出一般に成り立ち、要請されるのは層間の標本抽出が独立であることと、層毎に不偏推定量が得られていることのみである。各層について、 μ_h の不偏推定量 $\hat{\mu}_h$ と、その分散 $V(\hat{\mu}_h)$ 、分散の不偏推定量 $v(\hat{\mu}_h)$ が何らかの形で得られたとすると、これらを用いて、母平均の不偏推定量 $\hat{\mu}$ は

$$(5.1) \quad \hat{\mu} = \sum_{h=1}^L W_h \hat{\mu}_h$$

で与えられ、(5.1) の分散は

$$(5.2) \quad V(\hat{\mu}) = \sum_{h=1}^L W_h^2 V(\hat{\mu}_h),$$

分散の不偏推定量は

$$(5.3) \quad v(\hat{\mu}) = \sum_{h=1}^L W_h^2 v(\hat{\mu}_h)$$

で与えられる。

5.3 確率比例復元2段無作為抽出

第 h 層について、全部で M_h 個の第1次抽出単位 (PSU) から、その大きさに比例する確率で復元抽出 (PPR) し (重複分を含めて) m_h 個の PSU を得たとする。さらに抽出された第 j PSU ($j = 1, \dots, m_h$) が N_{jh} 個の第2次抽出単位 (SSU) を含むとして、そこから n_{jh} 個の SSU を非復元単純無作為抽出 (SI) で得るのが、(当該層における) 確率比例復元2段無作為抽出 PPR-SI である。通常は自動加重標本を得るために、全ての PSU の標本サイズを $n_{jh} = n_{0h}$ のように同じ大きさにする。

第 h 層の第 j PSU の第 i 番目の個体の特性値をあらためて y_{ijh} とすると、第 j PSU の平均値 μ_{jh} の不偏推定量 $\hat{\mu}_{jh}$ は、

$$\hat{\mu}_{jh} = \frac{1}{n_{jh}} \sum_{i=1}^{n_{jh}} y_{ijh}$$

で与えられ、これに基づき層の平均値 μ_h の PPR-SI 下での不偏推定量 $\hat{\mu}_{\text{PPR-SI}, h}$ は

$$(5.4) \quad \hat{\mu}_{\text{PPR-SI}, h} = \frac{1}{m_h} \sum_{j=1}^{m_h} \hat{\mu}_{jh}$$

で、その不偏な推定分散は、比較的簡単な形の

$$(5.5) \quad v(\hat{\mu}_{\text{PPR-SI}, h}) = \frac{1}{m_h(m_h - 1)} \sum_{j=1}^{m_h} (\hat{\mu}_{jh} - \hat{\mu}_{\text{PPR-SI}, h})^2$$

で与えられる (Cochran (1977), p. 308)。

5.4 層化確率比例復元2段無作為抽出

PPR-SI に層別を組み合わせた STPPR-SI 下での母平均の不偏推定量 $\hat{\mu}_{\text{STPPR-SI}}$ は、(5.1) の $\hat{\mu}_h$ として (5.4) を利用し

$$\begin{aligned}
 \hat{\mu}_{\text{STPPR-SI}} &= \sum_{h=1}^L \frac{W_h}{m_h} \sum_{j=1}^{m_h} \hat{\mu}_{jh} \\
 (5.6) \quad &= \sum_{h=1}^L \frac{W_h}{m_h} \sum_{j=1}^{m_h} \frac{1}{n_{jh}} \sum_{i=1}^{n_{jh}} y_{ijh} = \frac{1}{n} \sum_{h=1}^L \sum_{j=1}^{m_h} \sum_{i=1}^{n_{jh}} w_{ijh} y_{ijh}
 \end{aligned}$$

で与えられる。ここで、 $w_{ijh} = \frac{nW_h}{m_h n_{jh}}$ は標本の第 h 層第 j PSU に入る 1 要素のウェイトであり、 $n = \sum_{h=1}^L \sum_{j=1}^{m_h} n_{jh}$ は標本全体の大きさである。もし $w_{ijh} = 1$ であれば、

$$\hat{\mu}_{\text{STPPR-SI}} = \frac{1}{n} \sum_{h=1}^L \sum_{j=1}^{m_h} \sum_{i=1}^{n_{jh}} y_{ijh}$$

となり、自動加重となる。これを実現するには、 $n_{jh} = n_{0h}$ と層の中では PSU の標本サイズを一定にし、 $n_h = m_h n_{0h} = n W_h$ 、すなわち層の標本サイズの決定に比例割当を用いればよい。

不偏な推定分散は、(5.3) と (5.5) より、

$$(5.7) \quad v(\hat{\mu}_{\text{STPPR-SI}}) = \sum_{h=1}^L \frac{W_h^2}{m_h(m_h - 1)} \sum_{j=1}^{m_h} (\hat{\mu}_{jh} - \hat{\mu}_h)^2$$

で与えられる。この推定分散の平方根をとったものが (推定) 標準誤差 $\text{se}(\hat{\mu}_{\text{STPPR-SI}})$ となる。

本研究で扱うのは、もっぱら調査項目の回答肢の該当比率に関するデータである。母比率 p は個人がある属性に該当すれば $y_{ijh} = 1$ 、該当しなければ $y_{ijh} = 0$ のように 0-1 型のデータとした場合の平均値に他ならない。このときの $\hat{\mu}_{\text{STPPR-SI}}$ を特に $\hat{p}_{\text{STPPR-SI}}$ とかくことにする。

5.5 単純無作為抽出・デザイン効果

非復元単純無作為抽出 (SI) を仮定する場合、母平均の不偏推定量 $\hat{\mu}_{\text{SI}}$ は

$$(5.8) \quad \hat{\mu}_{\text{SI}} = \frac{1}{n} \sum_{i=1}^n y_i,$$

で、 $\hat{\mu}_{\text{SI}}$ の分散と不偏な推定分散は

$$(5.9) \quad V(\hat{\mu}_{\text{SI}}) = \frac{1-f}{n(N-1)} \sum_{i=1}^N (y_i - \mu)^2$$

$$(5.10) \quad = (1-f) \frac{\sigma^2}{n},$$

$$(5.11) \quad v(\hat{\mu}_{\text{SI}}) = \frac{1-f}{n(n-1)} \sum_{i=1}^n (y_i - \bar{y})^2$$

で与えられる。ここで、 σ^2 は母分散、 $f = n/N$ は抽出率である。これらの式は、 y_{ijh} を層と PSU を区別せず単に y_i (標本では $i = 1, \dots, n$ 、母集団では $i = 1, \dots, N$) とかいたものである。また本稿では、母分散 σ^2 として

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \mu)^2$$

のように母平均からの偏差の 2 乗和を N ではなく $N-1$ で除した、修正母分散 (鈴木・高橋 (1998) の用語) を用いている。

比率データでは母分散が母比率 p の関数として得られるため、(比率データの場合の $\hat{\mu}_{\text{SI}}$ を $\hat{p}_{\text{SI}}$ と書くと) 標本比率 $\hat{p}_{\text{SI}}$ の分散も p の関数として表され、次式となる：

$$(5.12) \quad V(\hat{p}_{SI}) = (1-f) \frac{N}{N-1} \frac{p(1-p)}{n}.$$

この分散の不偏推定量は

$$(5.13) \quad v(\hat{p}_{SI}) = (1-f) \frac{\hat{p}_{SI}(1-\hat{p}_{SI})}{(n-1)}$$

である。

(5.7) 式と (5.11) 式の比 $v(\hat{\mu}_{STPPR-SI})/v(\hat{\mu}_{SI})$ は、STPPR-SI の (推定された) デザイン効果 (Design Effect) と呼ばれるものである。以下、特定の標本抽出デザインの下でのある推定量の分散の、単純無作為抽出下での同じ母数の推定量の分散に対する比を Deff (推定量), その標本に基づく推定を deff (推定量) と略記し、 $\text{Deff}(\hat{p}_{PPR-SI})$, $\text{deff}(\hat{p}_{STPPR-SI})$ などと使う。本研究の場合比率データを扱い、層別 2 段抽出の下での母比率推定量 $\hat{p}_{STPPR-SI}$ の性質を議論するので、標本から deff を計算する際の分母には、 $\hat{p}_{STPPR-SI}$ と同じ母比率を持つ母集団から SI で得た推定量の分散の理論値として (5.12) 式の p に $\hat{p}_{STPPR-SI}$ を代入したものをを用いる。手元の同じ標本が単に SI から得られたものであると仮定して (5.13) 式により計算した値を分母に利用しても ($\hat{p}_{STPPR-SI}$ と $\hat{p}_{SI}$ が大きく異なる場合も多いので) 結果は殆ど変わらない。

本研究では、Deff の正の平方根をとった指標「標準誤差の倍率」を多く用いる。これを、 $\text{rSE}(\hat{p}_{STPPR-SI})$ ($\equiv \sqrt{\text{Deff}(\hat{p}_{STPPR-SI})}$) (Ratio of Standard Errors) のように表記する。標本に基づく推定には rse の略号を当てる。

5.6 異なる調査間でのデザイン効果と標準誤差の倍率の比較

異なる年次・標本設計下での標準誤差の比較を直接行うことは難しい。同一の調査項目だとしても、異なる年次の調査では標本サイズ n が異なる。どのような標本設計の仮定の下での推定にせよ、標準誤差の値には n が大きく影響しているから、 n が異なる調査間で標準誤差の推定値を単に比較しても得るところは少ない。この点については、Deff や rSE (あるいはその標本に基づく推定 deff, rse) は、ある程度 n のサイズの影響を抑えた指標になっているので、本稿では標準誤差の記述の中心的指標として rse を重用し、その分布の記述をやや丁寧に行う。

しかしこの指標も実は n と完全に無関係ではない。付録に示すように、STPPR-SI の下での Deff の大きさは、層別に関する相関比の 2 乗 (層間分散が全分散に占める割合) を ρ^2 , 2 段抽出に関する相関比の 2 乗 (PSU 間分散が層分散に占める割合: ただしそれが全ての層で一定であると仮定したもの) を η^2 として、概略次式で評価される:

$$(5.14) \quad \text{Deff}(\hat{\mu}_{STPPR-SI}) \simeq (1-\rho^2) \left\{ 1 + \left(\frac{n}{m} - 1 \right) \eta^2 \right\}.$$

ここで m は地点数である。(5.14) 式により、異なる調査間での Deff の比較に関する項は $((n/m)-1)\eta^2$ と $(1-\rho^2)$ の両方であるので、 n の違いの影響を評価する際には、 ρ^2 も η^2 も調査年次内で不変と仮定せざるを得ない。この仮定をおけば、 n の影響を大まかに調整することは可能であり、その調整の例を 6.4 節で簡単に示す。

さらに、本研究の文脈では、第 6 次・第 7 次と第 8 次以降の間で、層別の考え方に比較的大きな変更が加わった点も考慮する必要がある。すなわち第 8 次～第 10 次では、第 1 段の抽出において、一種の層別の効果を期待して層内での PSU の配列を工夫した系統抽出を行っていた。この標本設計の下では、第 6 次や第 7 次の設計下では層間分散の中に入っていた成分が、PSU 間分散に押し込まれる形になることで、 η^2 の相対的な増加と ρ^2 の減少を招くと考えられる。 η^2

の増加により各層での2段階抽出による推定量の分散が第6次・第7次に比べて相対的に過大評価される。更に $(1 - \rho^2)$ が1に近づき層別の効果が減殺される。これらが相乗的に作用し、結果としてrseなどの値も相対的に大きくなると予想される。

6. 標本誤差の比較結果

6.1 計算の内容と結果の例示

ここでは以下の記述の理解を助けるため、第6次から第10次の全ての年次で使用されている一つの質問項目を取り上げ、標本誤差の検討のために計算した指標の数値を例示する。取り上げた質問項目は、#4.10“他人の子供を養子にするか”で、これは、第1次(1953年)からの継続質問項目である。‘その他’と‘D.K.’を除き3つの選択肢(カテゴリ)が設定されており、カテゴリ2:‘つがせない’に関する計算結果を表6に示す。

表に示した指標は9種類である。[1]はいわゆる「単純集計」による比率、[2]と[3]はそれぞれ、標本設計が単純無作為抽出(SI)であると仮定した場合の[1]の推定分散と推定標準誤差である。また、[4]～[6]は5.4節で説明した層化2段階抽出STPPR-SIの仮定の下での計算であり、[4]が推定比率、[5][6]はその推定分散と推定標準誤差である。

残る3つは、SIの仮定下とSTPPR-SIの仮定下での数値の相違を評価する指標である。このうち[1]と[4]の差をとった[7]を本稿では「単純集計のバイアス2」と呼び、必要に応じBias2と略記する($\text{Bias2} = \hat{p}_{\text{SI}} - \hat{p}_{\text{STPPR-SI}}$)。自動加重標本が実現されていればこれは0になるが、計画標本段階で生じる端数による地点標本サイズの若干の不統一に加えて、回収標本では地点ごとの回収率の違いにより自動加重標本の前提が崩れるので、Bias2は自動加重標本の前提からの逸脱の程度を表す指標であると言える。表の第10次の例ではBias2は-1.825%となっており、単純集計と層化2段階抽出を仮定した重み付き集計では、2%弱の差が見られることになる。他の調査年次でも、これに近い値が得られている場合があるが、これは検討した第10次全カテゴリのBias2の分布の中でも絶対値が大き目のものである。本章6.2節では、この単純集計のバイアス2の分布を記述し、調査年次間で比較を行う。

また指標[8][9]は、それぞれ5.5節で定義したdeffとrseである。これらの指標及び計算の元になる[6]の値について、6.3節では、個々の調査における分布の検討を行い、6.4節では共

表6. #4.10“他人の子供を養子にするか”の回答肢‘つがせない’に関する標準誤差等の計算例。

計算した指標	式番号・意味等	第6次 K型 n = 2032	第7次 M型 n = 2173	第8次 K型 n = 1858	第9次 K型 n = 1833	第10次 K型 n = 1339
SIを仮定した計算						
[1] $\hat{p}_{\text{SI}}$		47.835	50.575	51.561	55.810	58.327
[2] $v(\hat{p}_{\text{SI}})$	(5.13)	1.229	1.151	1.345	1.346	1.817
[3] $se(\hat{p}_{\text{SI}})$	[2]の平方根	1.108	1.073	1.160	1.160	1.348
STPPR-SIを仮定した計算						
[4] $\hat{p}_{\text{STPPR-SI}}$	(5.6) の比率版	48.673	51.172	53.362	56.285	60.152
[5] $v(\hat{p}_{\text{STPPR-SI}})$	(5.7) の比率版	1.924	1.662	1.955	1.992	2.624
[6] $se(\hat{p}_{\text{STPPR-SI}})$	[5]の平方根	1.387	1.289	1.398	1.411	1.620
両者の違いの評価						
[7] $\hat{p}_{\text{SI}} - \hat{p}_{\text{STPPR-SI}}$	Bias2: [1] と [4] の差	-0.838	-0.596	-1.801	-0.475	-1.825
[8] $deff(\hat{p}_{\text{STPPR-SI}})$	本文中参照	1.564	1.445	1.459	1.483	1.465
[9] $rse(\hat{p}_{\text{STPPR-SI}})$	[8]の正の平方根	1.251	1.202	1.208	1.218	1.210

注1) [1][3][4][6][7]の単位は解り易さのために%に直してある。[2][5]の単位はその2乗。

通項目に基づく調査年次間の分布の比較を行う。

6.2 単純集計のバイアス2

本節では6.1節で定義した Bias2「単純集計のバイアス2」の大きさについて、年次間の比較を行う。

表7は、Bias1の場合と同様に（しかし今回の場合は性・年齢に関するカテゴリも含めて）バイアスの絶対値及びその対標準誤差比の分布を、各調査年次の調査票別に数値要約したもの、

表7. バイアス2の絶対値とその対標準誤差比の分布に関する数値要約(1)^{a)}.

	第6次		第7次		第8次		第9次		第10次	
	K型	M型	K型	M型	K型	M型	K型	M型	K型	M型
標本サイズ	2032	1913	2256	2173	1858	1824	1833	1905	1339	1341
カテゴリ数	132	145	163	194	164	168	172	219	179	211
絶対値の分布										
平均	0.257	0.248	0.267	0.199	0.601	0.498	0.418	0.344	0.444	0.425
標準偏差	0.237	0.191	0.258	0.184	0.551	0.451	0.397	0.345	0.398	0.361
最大値	1.376	0.978	1.461	1.082	3.446	2.949	2.059	2.350	1.825	1.728
90%ile	0.551	0.499	0.604	0.461	1.296	1.082	1.044	0.772	1.025	0.912
第3四分位	0.390	0.348	0.367	0.283	0.869	0.653	0.530	0.435	0.598	0.640
中央値	0.190	0.203	0.208	0.147	0.440	0.379	0.286	0.259	0.329	0.323
第1四分位	0.076	0.100	0.074	0.063	0.213	0.188	0.144	0.119	0.144	0.123
10%ile	0.035	0.048	0.038	0.029	0.079	0.067	0.058	0.047	0.077	0.063
最小値	0.003	0.003	0.004	0.000	0.000	0.001	0.003	0.000	0.006	0.002
対標準誤差比										
平均	0.289	0.280	0.318	0.241	0.632	0.538	0.448	0.383	0.413	0.392
標準偏差	0.234	0.208	0.276	0.194	0.520	0.451	0.397	0.361	0.330	0.315
1を超える割合(%)	0.8	0.0	3.1	0.5	20.1	13.7	10.5	5.5	6.1	3.8

注a) 各調査年次のK型, M型それぞれの有効カテゴリに基づいて計算した結果。単位%。

表8. バイアス2の絶対値とその対標準誤差比の分布に関する数値要約(2)^{a)}.

	第6次	第7次	第8次	第9次	第10次
絶対値の分布					
平均	0.268	0.263	0.599	0.385	0.502
標準偏差	0.243	0.205	0.462	0.343	0.455
最大値	1.376	0.775	1.801	1.482	1.825
90%ile	0.572	0.596	1.241	0.794	1.224
第3四分位	0.401	0.390	0.956	0.501	0.640
中央値	0.207	0.241	0.528	0.260	0.387
第1四分位	0.080	0.069	0.219	0.131	0.144
10%ile	0.038	0.039	0.079	0.066	0.060
最小値	0.003	0.011	0.000	0.021	0.006
対標準誤差比					
平均	0.290	0.290	0.595	0.387	0.445
標準偏差	0.246	0.213	0.422	0.316	0.375
1を超える割合(%)	1.2	0.0	7.1	2.4	6.0

注a) 5回の調査に共通の84カテゴリに基づいて計算した結果。

表8は同じ指標を5回の調査に共通の84カテゴリについて示したものである。表7に示したように、Bias2についても、最大で3.5%程度の値が得られる場合があるが、大略90%程度のカテゴリは絶対値が1%程度以下である(90パーセンタイルの値を参照)。ただしパーセンタイルの値は調査年次間でバラツキが見られ、それは特に第8次(1988年)で最大になっている。Abs(Bias2)の平均値も同様に第8次が最大で、以下第10次、第9次の順に大きく、第6次・第7次は同程度に小さいと見なし得よう。これに合わせるように、対標準誤差比が1を超える割合も第8次で最大になっており、平均値の大きさが一部の外れ値の影響によるとは考えにくい。

この数値の大小の順序は、共通項目のみに限って比較した表8でより明確に現れている。そしてこの大小の順序は表1の回収率の低さの順序と完全に一致する。すなわち、回収率が低い調査年次ほど、全般にAbs(Bias2)の値が大きくなる傾向が見られる。Bias2が(主として地点毎の回収率の違いによる)自動加重標本からの逸脱の程度の指標と見なしうることを考えると、回収率とのこの順序の一致は偶然とは考えにくい。回収率の低下は、全調査地点について一様に低くなるというよりも、地点毎の回収率自体のばらつきが大きくなる(回収が困難な地点と比較的良好な地点との差が出やすくなる)と考えられるからである。

なお、Bias2に関する5回の調査間の相関を共通の84カテゴリに関して求めると、第7次と第8次の間で例外的に0.660というやや高い値が得られた他は、ほぼ0.4以下の値であり、7.1.1で述べるBias1の場合ほどの一貫性は得られなかった。また異なる層別方法を用いた{第6次、第7次}と{第8次、第9次、第10次}の間の相関が弱くなるといった傾向も特に認められない。

6.3 個々の調査における精度の検討

6.3.1 標準誤差の大きさ

本節では各調査回各調査票毎に、標準誤差の大きさを示し、各調査票で大略どの程度の誤差幅を見込めば良いのかに関する具体的な数値を上げる。

標準誤差の大きさ自体は、比率の大きさによって意味が異なるから、全カテゴリについて代表値やパーセンタイル、あるいは散布度を求めることに積極的な意味はないが、参考のために

表9. 層化2段抽出を仮定して計算した標準誤差に関する数値要約^{a)}.

	第6次		第7次		第8次		第9次		第10次	
	K型	M型	K型	M型	K型	M型	K型	M型	K型	M型
標本サイズ	2032	1913	2256	2173	1858	1824	1833	1905	1339	1341
カテゴリ数	132	145	163	194	164	168	172	219	179	211
平均	0.979	1.001	0.921	0.908	1.099	1.088	1.094	1.065	1.228	1.249
標準偏差	0.283	0.273	0.255	0.277	0.306	0.302	0.313	0.303	0.333	0.322
最大値	1.467	1.395	1.383	1.445	1.576	1.627	1.539	1.550	1.698	1.699
90%ile	1.262	1.292	1.195	1.210	1.417	1.395	1.428	1.386	1.593	1.565
第3四分位	1.200	1.215	1.130	1.143	1.368	1.341	1.349	1.298	1.505	1.496
中央値	1.072	1.109	1.001	0.971	1.182	1.198	1.199	1.146	1.314	1.368
第1四分位	0.761	0.822	0.695	0.676	0.882	0.841	0.804	0.802	0.981	1.040
10%ile	0.554	0.578	0.547	0.474	0.640	0.591	0.609	0.562	0.702	0.731
最小値	0.328	0.324	0.358	0.316	0.341	0.346	0.401	0.351	0.423	0.388
標準誤差の目安 ^{b)}	1.240	1.256	1.160	1.195	1.386	1.356	1.386	1.337	1.549	1.539

注a) 各調査年次のK型、M型それぞれの有効カテゴリに基づいて計算した結果。単位%。

注b) 標準誤差の目安は、推定比率が35%以上65%以下のカテゴリの標準誤差の平均による。

表9にそれらの数値を示す。数値の単位はいずれも％に換算してある。

比率データの場合の標準誤差の値は、回答比率が50％程度のカテゴリで最大になるから、例えば推定比率が35％～65％程度のカテゴリの標準誤差を調べておけば、全カテゴリにわたる平均などよりもやや大きめの標準誤差の目安を与えることができる。表の最下段に示した「目安」の欄は、当該調査票で推定比率が35％～65％という条件を満たすカテゴリのみを取り出し、標準誤差の平均を求めたものである。こうして求めたものは各調査票のカテゴリでは第3四分位と90パーセンタイルの中間程度の値になっている。目安の値や90パーセンタイルの値などを見ると、例えば第10次であれば標準誤差として1.5％程度、95％信頼区間としてはこれを約2倍した $\pm 3\%$ 程度の誤差を見込んでおけば良さそうである、ということがわかる。あるいは第6次ではこれより幾分小さく、標準誤差にして1.25％程度、95％信頼区間としては $\pm 2.5\%$ 程度でよい。

表から明らかなように、標準誤差の大きさは同じ年次内のK型、M型の間では大略違いがなく、数値の違いは主として調査年次間に見られる。特に第10次の標準誤差の値が大きになっているが、これは標本サイズが小さいことが最も主要な理由である。標準誤差の目安など各種指標を見ても、標本サイズが大きい順に、標準誤差が小さくなっている。当然の結果ではあるが、標本サイズが小さい場合には、精度がやや落ちることを念頭に置いて集計数値を見る必要がでてくるということであり、この点は近年の国民性調査データの集計結果を見る際に押さえておくべき基本的事項と言えよう。

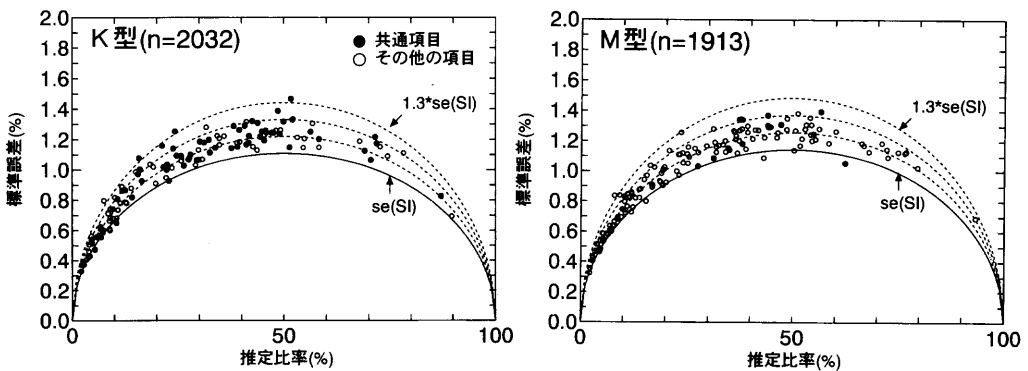


図3.1. 第6次調査項目の推定比率と推定標準誤差。

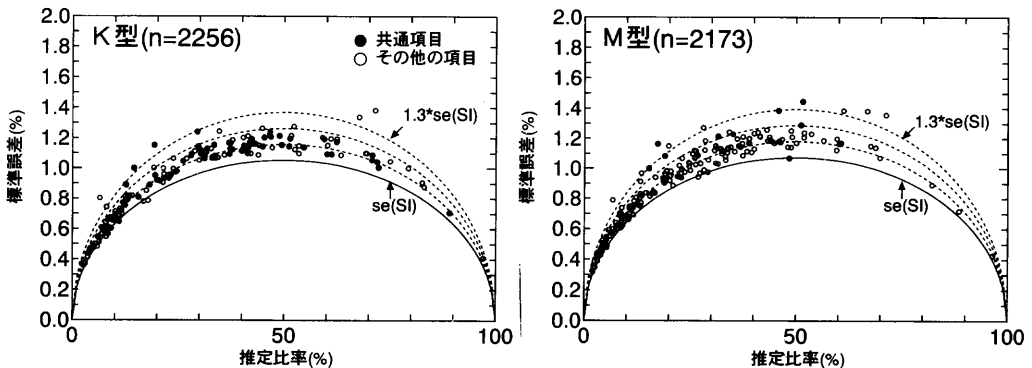


図3.2. 第7次調査項目の推定比率と推定標準誤差。

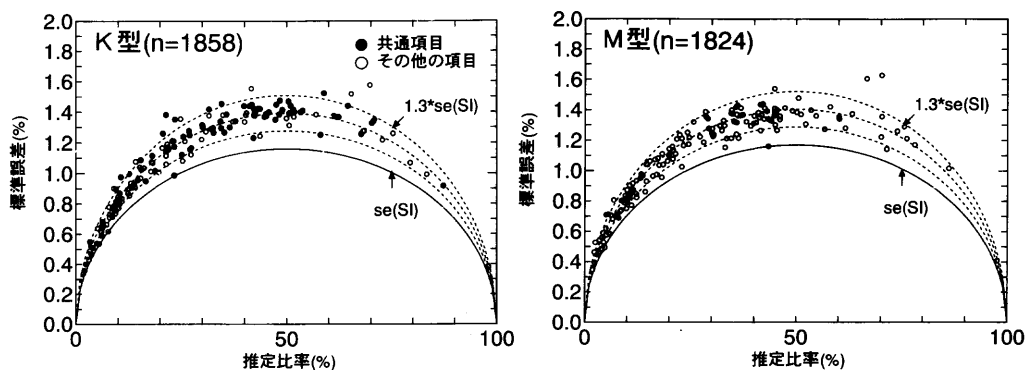


図 3.3. 第 8 次調査項目の推定比率と推定標準誤差.

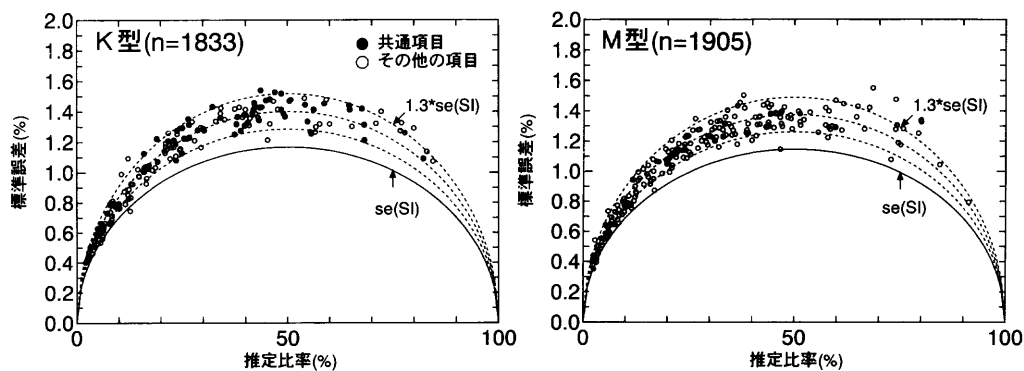


図 3.4. 第 9 次調査項目の推定比率と推定標準誤差.

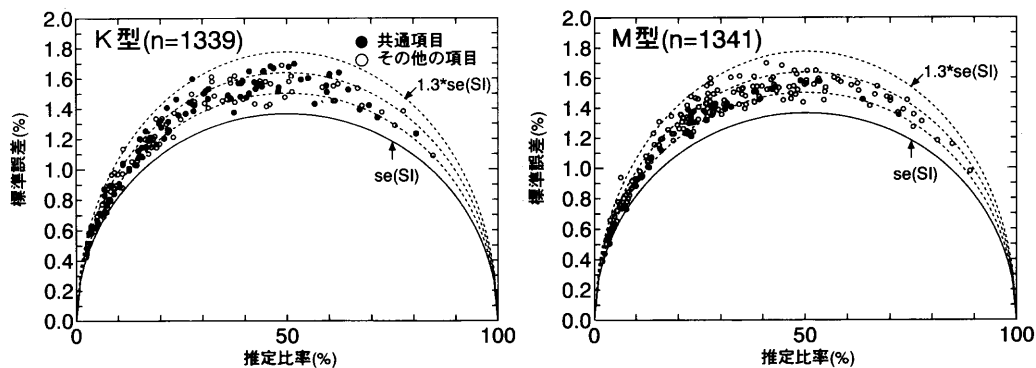


図 3.5. 第 10 次調査項目の推定比率と推定標準誤差.

6.3.2 標準誤差の倍率

図 3.1～図 3.5 は各回の調査の項目 (カテゴリ) の推定比率の (推定) 標準誤差を, 推定比率を横軸にしてプロットしたものである. 図の実線は SI を仮定した場合の標準誤差のラインであり, 点線は下から順にその 1.1, 1.2, 1.3 倍の値を示す. 縦軸のスケールを図の間で揃えて示してあり, 推定標準誤差の分布の代表値が異なるので, 各カテゴリの縦座標値の重心は図毎に少しずつずれていることになる. 図中に黒丸でプロットしたものは, 全 5 回の調査に共通だった 84 カテゴリである.

表 10. 標準誤差の倍率の分布に関する数値要約(1)^{a)}.

	第6次		第7次		第8次		第9次		第10次	
	K型	M型	K型	M型	K型	M型	K型	M型	K型	M型
標本サイズ	2032	1913	2256	2173	1858	1824	1833	1905	1339	1341
カテゴリ数	132	145	163	194	164	168	172	219	179	211
平均	1.115	1.103	1.118	1.117	1.194	1.185	1.189	1.180	1.141	1.148
標準偏差	0.082	0.070	0.081	0.081	0.089	0.092	0.085	0.090	0.070	0.069
最大値	1.345	1.314	1.507	1.459	1.522	1.595	1.444	1.557	1.337	1.467
90%ile	1.214	1.199	1.182	1.197	1.311	1.281	1.290	1.280	1.233	1.223
第3四分位	1.163	1.147	1.148	1.147	1.235	1.226	1.249	1.222	1.184	1.182
中央値	1.113	1.102	1.107	1.107	1.199	1.180	1.193	1.177	1.138	1.144
第1四分位	1.062	1.055	1.072	1.066	1.134	1.134	1.130	1.128	1.092	1.107
10%ile	1.019	1.016	1.045	1.028	1.082	1.080	1.081	1.071	1.052	1.071
最小値	0.926	0.946	0.952	0.987	0.967	0.957	0.934	0.959	0.957	0.924

注 a) 各調査年次のK型, M型それぞれの有効カテゴリに基づいて計算した結果.

表 11. 標準誤差の倍率の分布に関する数値要約(2)^{a)}.

	第6次	第7次	第8次	第9次	第10次
平均	1.121	1.129	1.212	1.197	1.152
標準偏差	0.080	0.083	0.075	0.069	0.061
最大値	1.331	1.432	1.448	1.330	1.305
90%ile	1.212	1.263	1.313	1.288	1.231
第3四分位	1.179	1.150	1.250	1.251	1.199
中央値	1.115	1.112	1.215	1.203	1.155
第1四分位	1.063	1.078	1.164	1.144	1.109
10%ile	1.035	1.059	1.117	1.106	1.080
最小値	0.941	0.971	1.002	1.040	0.975

注 a) 5回の調査に共通の84カテゴリに基づいて計算した結果.

一方標準誤差の倍率 rse の数値要約を K 型, M 型の調査票別に示すと, 表 10 の通りとなる. これらの図表から以下のことが指摘できよう.

- (1) 標準誤差の倍率が 1.3 倍を超える項目は少なく, また 1 を下回る項目もまれである.
- (2) 第 6 次と第 7 次はだいたい 1.1 倍から 1.2 倍の間に収まるのに対し, 第 8 次, 第 9 次は 1.2 倍から 1.3 倍の間に収まる. 第 10 次はその中間くらいの値のことが多い.
- (3) 全 5 回に共通のカテゴリについても他のカテゴリと特に異なる点はない.

多くの社会調査法のテキストに, 2 段階抽出では (SI に比べて標準誤差にして) 1.5 倍, 層化 2 段階抽出ならばその 1 ~ 2 割引き程度 (SI の 1.2 ~ 1.35 倍), の誤差を見込んでおけば良いといった目安が書かれているが (例えば杉山 (1984) など), 第 1 の指摘点はこうした記述と符合するものであると言えよう. この結果は, 近年の国民性調査の場合, 調査の年次が変わっても大同小異であることになる. 細かい差異については次節で検討する.

6.4 共通項目による調査間の標準誤差の比較

6.4.1 標準誤差の倍率の調査間での比較

全 5 回に共通の 84 カテゴリについて標準誤差の倍率 rse の数値要約を表 11 に示した.

この数値要約を見ると、{第6次, 第7次}, {第10次}, {第8次, 第9次}のように平均値をグループ化できそうである。第7次と第10次の間の差はややはっきりしないが、層別の方式が両者で異なっていることも考慮すれば、このグループ化はまず直観的には納得が行くであろう。

6.4.2 標準誤差の倍率の差の原因

以下標準誤差の倍率にこのような平均の違いが見られる原因を2点考察する。

第1点は、同じ層別方法下での結果である第8次・第9次と、第10次の結果間の差である。これは標本サイズ n の違いに帰することができる部分があると考えられる。5.6節で述べた通り、 rSE (というよりもその2乗である $Deff$) は n と完全に無関係ではない。他の条件が一定なら、 $Deff$ の評価式 (5.14) より n/m (地点当り標本サイズ) が小さいほど、小さくなる。各調査年次間で地点数 m の値は基本的に300に固定されている^{B17)}から、(5.14) 式で層の分離の程度を表す相関比の2乗 ρ^2 を固定して考えれば、層間で共通と仮定した2段抽出に関する相関比の2乗 η^2 の大きさを概算することが可能である。例えば rse の値が第10次の共通項目の平均程度である1.152の項目の場合、第10次の n をK型、M型回収標本サイズの平均をとって1340とし、 $\rho^2 = 0.1$ と仮定すると、

$$1.152^2 = (1 - \rho^2) \cdot \left\{ 1 + \left(\frac{n}{m} - 1 \right) \eta^2 \right\} = 0.9 \cdot \left\{ 1 + \left(\frac{1340}{300} - 1 \right) \eta^2 \right\}$$

を解いて $\eta^2 = 0.137$ となる。 η^2 と ρ^2 の値が不変で仮に標本サイズが第8次や第9次の回収標本並みの $n = 1850$ になったとすると、 rse の値は $\sqrt{0.9 \cdot (1 + ((1850/300) - 1) \times 0.137)} \simeq 1.240$ に引き伸ばされる。これは第8次の平均をやや上回る値である。

このような調整結果を考えると、標本サイズが同程度であれば、第10次の rse の平均は第8次や第9次の値と同程度かむしろやや大きくなると見積もられる。第10次において第8次、第9次よりも rse が小さいと見るのは数値の表面的な解釈に過ぎない。

以上の考察は相関比に関するいくつかの仮定に基づくものであり、細かい差異に捉われることは意味がないであろう。むしろ「第8次と第9次の間では回収標本サイズがだまかに似ているから、結果の数値もほぼ類似していると評価してよい。第10次から標本サイズの影響を取り除けば、第8次、第9次、第10次の rse の平均の差は殆ど取るに足らない」と考えておくことにする。

とすれば、考察すべき第2の無視できない差は {第6次, 第7次} と {第8次, 第9次, 第10次} の間に存在すると考えられる。第6次, 第7次のほうが地点当りの標本サイズが大きい。標本サイズが大きいことは rse を大きくする方向に寄与するから、両グループの差は標本サイズの差によって説明されるものではない。したがって少なくとも差の一部は、層別の方法を変更したことに伴う、後者のグループでのPSU間分散の(相対的な)過大評価(5.6節)に基づくものであることが予想される。しかしながら、詳細は省略するが、試みに第8次～第10次の実際の層別の枠を取り払って、第7次と類似した層が得られるように調査地点を60層程度に(事後)層別し、STPPR-SIの仮定の下で計算を行うなどしても、推定標準誤差あるいは rse の縮小について期待されるような結果は得られず、上記の予想を積極的に裏付ける証拠は得られなかった。よって、{第6次, 第7次} と {第8次, 第9次, 第10次} 間の差の原因については、今後詳細に検討する必要がある。

7. 考 察

7.1 バイアスの性質及び関連する事項について

本研究では性質の異なる2種類のバイアス (Bias1, Bias2) について調査年次間の比較を行った。バイアスとはともに、単純集計による比率 $\bar{p}_{SI}$ の、重み付き集計による比率からの差で定義した。国民性調査の結果は従来、単純集計数値として発表されており、この数値のある種の「頑健さ」を評価したものになっている。

本節では2種類のバイアスの性格と、それに基づく検討結果に関連して若干の事項を補足する。

7.1.1 バイアス1の性質及びそれを導出した重み付き集計値について

バイアス1は、性・年齢で回収標本を事後層別し、母集団の層のサイズで重み付けた集計値 $\bar{p}_{PST}$ を求め、 $Bias1 = \bar{p}_{SI} - \bar{p}_{PST}$ としたものであった。これは回収標本の性・年齢構成の母集団のそれからの乖離が第9次・第10次調査で顕著になっていることの影響を見るために算出したものである。

4.3節ではバイアス1を単純集計数値の(真値からの)偏りの目安として使用したが、 $\bar{p}_{PST}$ が常に良い推定であるとは言い切れない。なぜなら、 $\bar{p}_{PST}$ は調査可能であった者の回答のみを集計することになるため、調査協力者の属性や回答を過大評価する——若年層での回収率が低いことを考慮すると、特に若年層の協力者のそれを過大評価する——という別のバイアスを持ち込む恐れがあるからである。

社会調査データで回収不能による欠測を補間し真値を適切に推定することは、方法の提案がいくつかあるものの、基本的には難しい問題である。 $\bar{p}_{PST}$ も欠測を補間する素朴な方法の一種と見なしうが、上述の理由で、本稿は $\bar{p}_{PST}$ を推奨する立場でこれを利用したものではない。現実にはバイアス1の値が多く項目で小さいこと、費用対効果のことを考えると、国民性調査の場合に「事後層別した重み付き推定を単純集計の代わりに用いるほうが良い」と、簡単に結論付けることはできない。

他方、性・年齢に関する補正に限って言えば、母集団における性・年齢構成情報(層の重み)は一般に入手しやすい場合が多いので、個人を対象とするどのような調査にとっても $\bar{p}_{PST}$ は計算が容易であるという利点を持っていると言える。

また、本研究では $\bar{p}_{PST}$ から調査項目ごとに計算されるバイアス1を調査間で比較することによって、バイアス1が各回の調査を特徴づける一つの指標になっていることは確認できる。例えば、5回の調査に共通の84カテゴリ(これらには属性項目は含まれていない)について Bias1 の調査年次間の相関を計算すると、最低でも第9次と第10次の間の0.381、最大は第6次と第7次の間で0.916、5回の調査の組み合わせ ${}^5C_2 = 10$ 個の相関係数の単純平均は0.641となる。つまり、バイアス1の大きさは調査年次を通じて一貫性がかなり高く、このことは項目の性質を予め承知しておけば、将来の調査でも $Abs(Bias1)$ が大きくなるか、あるいは無視し得る程度であるかの予想がつくことを意味している。このような観点から、バイアス1の利用法を考えることも可能である。

回収率の問題は今後劇的な改善を期待することはできず、現在の調査方式を続ける限り性・年齢構成の母集団分布からの乖離は今後も続くものと予想できる。性、年齢と連関の強い項目がそうした乖離の影響を受けやすいことは自明であるが、標本設計あるいは集計面を含む調査方式における対策を考えるためには、上記のような乖離に対する単純集計 $\bar{p}_{SI}$ の頑健さをより詳細に評価する必要がある。項目自体の性・年齢との連関の程度や、属性と連関する欠測(不能票)がどの程度の割合で生じるかの水準、等を要因とした数値実験を行うことが有効であろう。

今後の検討課題としたい。

7.1.2 バイアス2の具体的意味と評価

バイアス2は $\text{Bias2} = \bar{p}_{\text{SI}} - \bar{p}_{\text{STPPR-SI}}$ によって定義したが、これが何の指標であるかを端的に言語化することは難しい。自動加重標本が達成されている場合に0となる指標であり、自動加重の前提からの逸脱の程度を表しているとは言えるが、同じ調査年次の同じ標本内の項目でも、逸脱の影響の大きさに違いがある。どういう項目で逸脱の影響が大きくなり、正方向と負方向のバイアスがそれぞれ何を意味するのかについては、項目の性質毎に考察しなければならない。バイアス2が大きくなる理由を類型化することは、調査項目の設計の上でも必要なことと考えられるので、今後の検討課題とする。

バイアス2が大きくなる理由が明確でないので以下の考察も暫定的な内容ではあるが、回収率の低下が、バイアス2には影響を与えたように見える。6.2節に述べたように地点毎の回収率のバラツキが原因となって、回収率の低い年次ほど平均的に大きなバイアスが得られたと推論可能である。換言すると、回収率が低い場合 STPPR-SI を仮定した本来の推定量との差が大きくなるという意味で、単純集計が妥当でなくなる恐れがある。単純集計数値の自動加重標本の前提からの逸脱に対する頑健さについて精査するためにも、数値実験は有効であろう。

7.2 標準誤差と標準誤差の倍率について

以下では、 Deff と rSE に関して、本研究で残された課題を中心に言及する。

7.2.1 Deff の規定因の検討

まず(5.14)式と付録より、同じ調査の中では、

- ・地域的に偏在する性質を持つために PSU 間分散が大きくなり、2段抽出の影響が大きくなる(η^2 の大きさによって Deff_0 が大きくなる)項目
- ・層別の効果が出にくい(ρ^2 が小さい)項目

において $\text{deff}(\bar{p}_{\text{STPPR-SI}})$ や $\text{rse}(\bar{p}_{\text{STPPR-SI}})$ が大きくなりやすいと言える。層別2段抽出においては、同じ層内では性格の似た PSU が集まっていると期待されるが、2段抽出に関する相関比の2乗 η^2 と層別に関する相関比の2乗 ρ^2 の関係は、一方が大きい場合に他方は常に小さいといった単純なものではなく、両者の力関係は項目の性格により様々であることが予想される。詳細は示さないが、標準誤差の倍率について調査年次間で相関係数を求めると、多くの組み合わせで0.3~0.5程度であり、ある程度の一貫性を有することが判っている。各カテゴリ毎に、 η^2 あるいは層毎の2段抽出に関する相関比の2乗 η^2 と、層別に関する相関比の2乗 ρ^2 の具体的な推定を行うことにより、当該項目の rse の大小の原因やその一貫性・安定性に関する知見が得られると期待できる。今後の検討課題とする。

7.2.2 異なる標本設計下での比較

ところで rSE と Deff を5回の調査の間で比較してきたが「標本サイズが異なる」「層別方法が異なる」といった、標本設計の条件の異なる調査間では、その数値の表面上の差を単純に標本精度の良し悪しの評価に結びつけることは控えるべきである。数値の大きさに影響を与える要因も2種類の相関比と地点当たり標本数が絡んでおり、やや複雑であった。こうした困難を抱える中でも、 rse に関する第8次、第9次の結果に断層と言えるような差がないことは、この2回については調査全体としての(標準誤差に関係する要因についての)精度は同程度であったことを示唆している。この2回では、(回収)標本サイズが似ているばかりではなく、層別の

方式(「はじめに」で述べた条件2)も同じであるから、rSEを比較しやすい条件が揃っている。他方この第8次と第9次の間に「実査担当機関の違い」(条件1)と比較的大きな「回収率の差」(条件3)があった点を合わせると、実査担当機関の違いや回収率の差は、少なくとも層別方式の違いを超えるような推定量のばらつきの差を生み出す要因にはなっていないと推論可能である。

一方で第6次・第7次と第8次以降の間に「断層」と表現できそうな大きな差があった点についての理由を説明できないという課題が本研究では残された。確かに5.6節の予想通り、{第8次, 第9次, 第10次}では{第6次, 第7次}に比べて相対的に大きなrseの値が得られた(6.4節)。5.6節では「第6次・第7次では層間分散に組み込まれた成分が、第8次～第10次の場合にはPSU間分散の側に押し込まれる形になる、したがってPSU間分散の相対的過大評価と層間分散の相対的な減少が重なり、Deff, rSEの相対的過大評価を招く」との解釈を示した。しかし差の原因がそれだけであるのか、それとも別の原因もあるのかは不明である。例えば表1に示したように、第8次以降は層別方式が変わっただけでなく第6次・第7次比べて回収率が低いという条件も抱えている。これらの影響が全くなかったと言い切ることはできない。この課題についても、二種類の相関比の2乗(ρ^2 と η^2)の値の具体的な推定が一つの見通しを与えると考えられる。

7.3 本研究で得られた知見の要約

以下に本研究で得られた知見を要約し、結語の代わりとする。

本研究では、第6次から第10次の5回にわたる「日本人の国民性調査」全国調査の標本について、性・年齢構成と標本誤差などを検討し、以下の結果を得た。

- 1) 回収標本の性・年齢構成は、計画標本と比べて母集団の性・年齢構成からの乖離が大きく、その乖離の程度は第9次と第10次で大きくなってきている。さらに、乖離は性・年齢に関してランダムに生じるのではなく、特定の傾向、すなわち特に男性を中心とする若年層の比率が小さく、それを補うように、女性中高年層(40歳代以降)と男性高年層(60歳代以降)の比率が大きくなるという方向である。
- 2) 上記1)の影響により、単純集計結果は真値に比べて、高年齢層の属性や回答の方向に偏った数値になっていると推測できる。事後層別の方法で簡便にこの偏りを評価した結果によれば、単純集計の偏りは多くの項目で1%程度以内であること、第10次の結果ではこの偏りが、従来の結果と比べて異質に大きいこと、の2点を指摘できる。また偏りの調査年次間の相関は高く、偏りが大きい項目は調査間で一貫して大きい傾向が見られた。
- 3) 各年次の標本設計が層化確率比例復元2段無作為抽出(STPPR-SI)であると仮定して、ほとんど全ての調査項目のカテゴリに関して、母比率推定量の標準誤差を計算したところ、第10次では標本サイズの小ささのために、相対的には他の年次に比べて大きな値が得られた。95%信頼区間の幅をやや大きめに推定しておく、第10次では多くの項目で $\pm 3\%$ を見込めばよい。他の年次ではそれより少し小さいが、それでも $\pm 2.5\%$ 程度を見込む必要がある。
- 4) 上記3)で求めた標準誤差の、単純無作為抽出(SI)を仮定した場合の標準誤差に対する倍率は、その値が小さめである第6次と第7次では1.1～1.2倍を中心に分布し、やや大きめである第8次と第9次では、1.2～1.3倍を中心に分布していた。第10次の場合も標本サイズの影響を除けば、第8次、第9次と同程度か、それよりやや大きい程度であると推定される。これらはいずれにせよ従来から誤差の倍率に関して文献類に示されている目安の数値の範囲内である。

- 5) SI を仮定した単純集計と STPPR-SI を仮定した加重集計との差を検討したところ、差の大きさは回収率の低さと関係することが示唆された。5 回の中で回収率が最低の第 8 次調査で両集計値の差が最大になり、以下回収率の低い順に、第 10 次、第 9 次などとなった。しかし、この数値も絶対値としては 1%未満の場合が多く、従来から単純集計により発表されている国民性調査の結果数値から大きくずれるようなものではない。

注.

- 注 1) 本稿では「標本誤差」の語を、標準誤差（あるいは分散）という推定量のブレの大きさによって評価される標本抽出に伴う誤差を指すのに限定して用いる。その他に結果数値に影響を与える誤差要因を「非標本誤差」と考える。「標本精度」は、標本誤差と非標本誤差の大きさを踏まえた上での結果の数値の正確さを指す語として用いる。したがって、本稿の標題には、非標本誤差と標本誤差の両者を検討することを含意させている。前者については具体的には標本の性・年齢構成の偏りという一つの側面のみについて検討する。本稿の目的は結果を総合して調査精度に評価を与えるというよりも、それぞれの側面についての基礎的検討資料を提示すること自体にある。
- 注 2) ここでいう調査方式とは、調査員による個別面接聴取法によるということ、調査票の具体的な構成、実査の進め方、対象者のサンプリング方法（層化多段無作為抽出の具体的な設計方法）のことなどを含む調査方法の総体を指す。
- 注 3) 国民性調査は、同一質問項目の継続利用を基本としてはいるが、各回の調査項目の組が完全に同一であるわけではなく、その点で質問項目の配列を含む調査票の構成も、各回でいくらかは異なることになる。例えばそういった細かい実査の条件も、ここでは調査方式に含めて考えるわけである。
- 注 4) 実査担当機関の違いがもたらす一つの明白な調査条件の差異は、統計数理研究所による実査では、調査員がアルバイトの学生であるのに対し、専門調査機関への委託では、調査員は（必ずしも専従者とは限らないが）普段から調査員としての活動をしている専門家である、という点である。実査担当機関の違いがそれ以外にどのような違いを含み、具体的にどのような実査プロセスの差をもたらしているかを全て指摘することは難しい。なお、集計数値に基づく考察については坂元 (1995 a, 1995 b) 参照。
- 注 5) 異なる調査間では調査時期という要因が違っているわけであるから、この意味で本研究で行う比較の意味は、剰余変数を統制した要因実験（実験調査）による結果の比較とは、異ならざるを得ない。すなわち、数値の違いの原因を条件の違いに直接帰することはできないことになる。
- 注 6) 表 1 を一見すると分かるように、第 7 次と第 8 次の間で回収率に大きな差があり、これは条件 2 の層別の方法の変更と同時に起きている。しかし、層別方法の変更が、特別に調査困難地点の増加を招いたとは考えられず、この回収率の低下はむしろ良質の大学生調査員の調達が困難になったという、第 8 次調査時の実査担当機関の側の問題であると考えられる。実際第 9 次調査では同じ層別方法の下で、回収率が 69%にまで回復していることが、層別の方法と回収率は独立であるとの考察を裏付けている。
- 注 7) SSM 調査は、社会学者が調査委員会を構成して 1955 年以來 10 年おきに実施している継続学術調査である。同様の時期に開始された長期にわたる継続調査であることの他に、標本設計も日本人の国民性調査と毎回ほぼ同じ枠組で行われており、二つの調査は双子の兄弟のような存在である。統計数理研究所のスタッフも 1965 年の第 2 次調査以来、毎回 SSM 調査委員会に加わっている。
- 注 8) 投票区別の有権者数調べにあたっては、各調査年次とも毎日新聞社の世論・選挙センター

- (旧世論調査部)のご協力を得た。記してあらためて感謝の意を表する次第である。
- 注9) ここで自動加重標本とは、(重み付き集計でなく最終抽出単位の重みを等しくした)単純集計が母集団平均値の不偏推定量を与えるように設計された標本のことである。
- 注10) 後の確率比例2段階抽出を仮定した標本誤差の計算のために、層内の地点数が2以上である必要がある。このため第7次で層内地点数が1であった層を他の層と併合し、実質的に第7次は55層として計算を進める。
- 注11) 実際、第6次の地点毎の計画標本サイズは、15~20(地点当り平均18)、第7次のそれは17~22(地点当り平均20)と、第8次以降と比べると最大と最小の差がやや大きい。第8次以降の方式では沖縄の2地点を除いて差が高々2となる。
- 注12) 乱数を操作する際の手続きとしては、「小数部分を切り捨てた数値に1を加える」とするのが正確な表現である。
- 注13) バイアスという語は、通常「推定量の期待値と真値との隔たり」の意味で使われる(真値とは本稿の文脈では母比率である)。本稿では他に適切な語を見出し得なかったため、仮にバイアスという語を用いている。この用語法によって、加重集計の側が真値であることを主張するものではない。
- 注14) 標本サイズの影響を除くには、バイアス1の2乗を推定分散で割るのが適切であるが、ここで考えた指標はその正の平方根であり、直観的な意味はこちらの方が分かりやすい。どちらの指標で考えても導かれる結論に相違はない。
- 注15) 最終抽出単位の抽出が系統(等間隔)抽出になっていることを考慮した標本誤差評価の試みが、統計数理研究所国民性調査委員会(1975), pp. 417-423 でなされている。本研究では、この問題についてはこれ以上言及せず、2段階目の抽出についてはSIを仮定した議論のみを続ける。
- 注16) この仮定[5]が強い仮定であるというのは、回収の有効・無効は、第4章に見るように例えば年齢という属性とは関係が深いというのが、この仮定に反する一つの証拠であるし、あるいは調査不能の多くの部分を占める一時不在と拒否による不能が、質問項目が測定しようとしている態度と独立であるというのも考えにくい、なども論拠となり得よう。しかしながら、回収不能が標本精度に与える影響を評価する方法、あるいは回収不能を適切に補正する方法についてはにわかな解決を見出しがたく、本研究では、回収標本には既に述べたような偏りがあることを承知しつつ、回収標本があたかも当初からの計画サンプルであるかのように扱い、標本誤差の計算を行う。
- 注17) 厳密に言えば、計画標本段階では $m = 300$ であるが、まれに当該地点でK型あるいはM型の調査票対象者全てが調査不能になることがあり、回収標本では $m = 300$ より若干減るケースもある。

付録：デザイン効果と標準誤差の倍率の意味

ここでは本研究で標本誤差の指標として重用するデザイン効果(Deff)と標準誤差の倍率(rSE)の性質について説明する。後者は前者の平方根であるから、前者の性質を調べる。

A.1 確率比例復元2段階抽出におけるデザイン効果

まず層別を行わない確率比例復元2段階抽出(PPR-SI)と単純無作為抽出(SI)の比較を考える。各層ごとにPPR-SIを行うSTPPR-SIのケースに備えて、層を表す添字 h を付けた式として与える。

層毎に行ったSIにおける層の平均 μ_h の推定量 $\hat{\mu}_{SI,h}$ の分散は、層の標本サイズ n_h 、抽出率 f_h

と層分散 σ_h^2 の関数として (5.10) 式に添字 h を付けた次式により表される：

$$V(\hat{\mu}_{\text{SI},h}) = (1 - f_h) \frac{\sigma_h^2}{n_h}.$$

全国調査などでは $(1 - f_h)$ の項が無視できるから、実質的に $V(\hat{\mu}_{\text{SI},h}) \simeq \sigma_h^2/n_h$ で評価される。ここで層分散 σ_h^2 は次式で定義されるものである：

$$(A.1) \quad \sigma_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{ih} - \mu_h)^2 = \frac{1}{N_h - 1} \sum_{j=1}^{M_h} \sum_{i=1}^{N_{jh}} (y_{ijh} - \mu_h)^2.$$

2 段抽出であることを明示するために h 以外に 2 つの添字を用いた表記が最右辺である。 y_{ijh} は第 h 層の第 j 番目の PSU における第 i 番目の要素の特性値である。

層分散 σ_h^2 は次のように PSU 間分散と PSU 内分散の和に分解される：

$$\begin{aligned} \sigma_h^2 &= \sigma_{b,h}^2 + \sigma_{w,h}^2, \\ \sigma_{w,h}^2 &= \sum_{j=1}^{M_h} \frac{N_{jh} - 1}{N_h - 1} \sigma_{jh}^2 \quad (\text{PSU 内分散}), \\ \sigma_{b,h}^2 &= \sum_{j=1}^{M_h} \frac{N_{jh}}{N_h - 1} (\mu_{jh} - \mu_h)^2 \quad (\text{PSU 間分散}). \end{aligned}$$

この性質を用いて、層毎の SI における分散評価式を PPR-SI の場合の分散評価式に合わせる形で表現すると

$$(A.2) \quad V(\hat{\mu}_{\text{SI},h}) \simeq \frac{\sigma_h^2}{n_h} = \frac{1}{n_h} (\sigma_{w,h}^2 + \sigma_{b,h}^2)$$

となる。

一方 PPR-SI における層平均の不偏推定量 (5.4) の分散の評価にも PSU 間分散と PSU 内分散が関与し、次式の形となる：

$$V(\hat{\mu}_{\text{PPR-SI},h}) = \frac{1}{n_h} \sum_{j=1}^{M_h} \frac{N_h - n_{jh}}{N_h} \sigma_{jh}^2 + \frac{1}{m_h} \sum_{j=1}^{M_h} \frac{N_{jh}}{N_h} (\mu_{jh} - \mu_h)^2.$$

右辺第 1 項の n_{jh}/N_h (PSU からの SSU の抽出率) を 0 と見なし、 $(N_h - 1)/N_h \simeq 1$ といった近似を適宜利用して、上式は

$$(A.3) \quad V(\hat{\mu}_{\text{PPR-SI},h}) \simeq \frac{1}{n_h} \sigma_{w,h}^2 + \frac{1}{m_h} \sigma_{b,h}^2$$

のように変形される。これが $V(\hat{\mu}_{\text{PPR-SI},h})$ の近似的評価である。

層毎に行った PPR-SI のデザイン効果は、(A.3) 式と (A.2) 式の比であるから

$$\begin{aligned} (A.4) \quad \text{Deff}_h &\equiv \text{Deff}(\hat{\mu}_{\text{PPR-SI},h}) \equiv \frac{V(\hat{\mu}_{\text{PPR-SI},h})}{V(\hat{\mu}_{\text{SI},h})} \\ &\simeq \frac{\frac{1}{n_h} \sigma_{w,h}^2 + \frac{1}{m_h} \sigma_{b,h}^2}{\frac{1}{n_h} \sigma_{w,h}^2 + \frac{1}{n_h} \sigma_{b,h}^2} = 1 + \left(\frac{n_h}{m_h} - 1 \right) \frac{\sigma_{b,h}^2}{\sigma_h^2} \\ (A.5) \quad &= 1 + \left(\frac{n_h}{m_h} - 1 \right) \eta_h^2 \end{aligned}$$

となる。ここで η_h^2 は (第 h 層における PSU 間分散 $\sigma_{b,h}^2$ の層分散に対する寄与の程度を表す) 相関比の 2 乗である。この正の平方根 η_h を 2 段抽出に関する相関比と呼ぶ。既述の標本設計から、少なくとも計画標本段階では $(n_h/m_h) - 1$ (すなわち第 h 層における (地点あたりの標本サイズ - 1)) は (特に第 8 次～第 10 次では) 層を通じて各調査毎にほぼ一定になっており、 Deff_h の主要な構成要素は $\sigma_{b,h}^2/\sigma_h^2 = \eta_h^2$ の部分であることが判る。つまり PSU 間分散の層分散に占める相対的な割合 η_h^2 が高ければ、 Deff_h が大きくなる。

簡単のために、以下では各層での Deff_h が $\text{Deff}_h = \text{Deff}_0$ ($h = 1, \dots, L$) の様に層間で等しいことを仮定する。 n_h/m_h が一定であれば、これは 2 段抽出に関する相関比の 2 乗が $\eta_h^2 = \eta_0^2$ と層間で共通であると仮定したことを意味する。 n_h/m_h が一定との仮定は本研究の標本設計では (計画標本の段階で近似的に) 成立しており、層間で共通に n/m と書くことができる。つまり、全ての層の Deff ($\hat{\mu}_{\text{PPR-SI},h}$) が

$$\text{Deff}_0 = 1 + \left(\frac{n}{m} - 1 \right) \eta_0^2$$

を仮定したことになる。

(A.4) 式を書き換えた次式は後に用いる：

$$(A.6) \quad V(\hat{\mu}_{\text{PPR-SI},h}) = \text{Deff}_h \cdot V(\hat{\mu}_{\text{SI},h}) \simeq \text{Deff}_h \frac{\sigma_h^2}{n_h}.$$

A.2 層別確率比例復元 2 段抽出におけるデザイン効果

以上は層ごとに議論を展開したが、母集団に関して PPR-SI を行うと上記から添字 h がとれた式が成立する。層別の効果が加わることによって、STPPR-SI では母平均推定量の分散が (層別を行わなかった場合に比べて) 小さくなる (したがって Deff の絶対値も小さくなる) ことが予想される。この事情は以下のように説明することができる。

まず、層別抽出における総分散の層内分散と層間分散への分解を (2 段抽出におけるそれと区別するために) 以下のように表記しておく：

$$(A.7) \quad \sigma^2 = \sigma_w^2 + \sigma_b^2,$$

ただし

$$\sigma_w^2 = \sum_{h=1}^L \frac{N_h - 1}{N - 1} \sigma_h^2 \quad (\text{層内分散}),$$

$$\sigma_b^2 = \sum_{h=1}^L \frac{N_h}{N - 1} (\mu_h - \mu)^2 \quad (\text{層間分散}).$$

層内分散の層分散に対する比 σ_b^2/σ^2 は層の分離の程度を示す指標である。これを 2 段抽出に関する相関比と区別して ρ^2 と書き、層別に関する相関比の 2 乗と呼ぶ。(A.7) を σ_w^2 について解くと、層内分散を

$$(A.8) \quad \sigma_w^2 = (1 - \rho^2) \sigma^2$$

という形で書くこともできる。比例割当による層別無作為抽出 STSI における母平均推定量 $\hat{\mu}_{\text{STSI}} = \sum_{h=1}^L W_h \hat{\mu}_{\text{SI},h}$ の分散が、近似的に $V(\hat{\mu}_{\text{STSI}}) \simeq \sigma_w^2/n$ で評価され、単純無作為抽出 SI での母平均推定量の分散 $V(\hat{\mu}_{\text{SI}}) \simeq \sigma^2/n$ に対する分散縮小率がほぼ $(1 - \rho^2)$ であること、つまり $(1 - \rho^2)$ が層別の効果となることはよく知られる結果である。

さて、STPPR-SI における母平均の不偏推定量とその分散は、(5.1) 式と (5.2) 式の $\hat{\mu}_h$ に

$\hat{\mu}_{PPR-SI,h}$ を利用したものになる。このうち分散については、比例割当では $W_h = N_h/N = n_h/n$ となることと (A.6) による近似を用いると、

$$\begin{aligned} V(\hat{\mu}_{STPPR-SI}) &= \sum_{h=1}^L W_h^2 V(\hat{\mu}_{PPR-SI,h}) \simeq \sum_{h=1}^L \frac{N_h}{N} \frac{n_h}{n} \cdot \text{Deff}_h \frac{\sigma_h^2}{n_h} = \text{Deff}_0 \frac{1}{n} \sum_{h=1}^L \frac{N_h}{N} \sigma_h^2 \\ &\simeq \text{Deff}_0 \frac{1}{n} \sum_{h=1}^L \frac{N_h}{N-1} \sigma_h^2 \simeq \text{Deff}_0 \frac{\sigma_w^2}{n} \end{aligned}$$

と評価される。(A.8) 式を考慮すると上式は、

$$(A.9) \quad V(\hat{\mu}_{STPPR-SI}) \simeq \text{Deff}_0 \cdot (1 - \rho^2) \frac{\sigma^2}{n}$$

を意味することになる。

以上から STPPR-SI におけるデザイン効果は、(A.9) 式と $V(\hat{\mu}_{SI}) \simeq \sigma^2/n$ の比をとって

$$\begin{aligned} (A.10) \quad \text{Deff}(\hat{\mu}_{STPPR-SI}) &\equiv \frac{V(\hat{\mu}_{STPPR-SI})}{V(\hat{\mu}_{SI})} \\ &\simeq \frac{(1 - \rho^2) \cdot \text{Deff}_0 \cdot \sigma^2/n}{\sigma^2/n} \\ &= (1 - \rho^2) \cdot \text{Deff}_0 = (1 - \rho^2) \left\{ 1 + \left(\frac{n}{m} - 1 \right) \eta_0^2 \right\} \end{aligned}$$

となる。STPPR-SI と PPR-SI の比較においても(層間で η_h^2 が一定という強い仮定が必要ではあるが)層別の効果として $(1 - \rho^2)$ 分の分散縮小率が得られる。層間で η_h^2 が一定ではない場合の式はこのようには簡単にならないが、層別の効果によって類似の分散縮小が得られているものと想定し得る。

参 考 文 献

- Cochran, W.G. (1977). *Sampling Techniques*, Third Edition, Wiley, New York.
- 中村 隆 (1998). 1995 年 SSM 調査の標本設計と標本精度 — 標本抽出法を考慮した分析に向けて —, 『社会階層・移動の基礎分析と国際比較』(石田 浩 編), 1995 年 SSM 調査シリーズ 1, 77-100, 1995 年 SSM 調査委員会, 東京大学・東北大学, 東京・仙台.
- 中村 隆, 土屋隆裕, 前田忠彦, 坂元慶行 (1999). 国民性の研究 第 10 回全国調査 — 1998 年全国調査, 統計数理研究所研究リポート, No. 83.
- 坂元慶行 (1995a). 意識調査の調査方式と調査結果その①, *ESTRELA*, 1995 年 10 月号, 43-51, (財)統計情報研究開発センター.
- 坂元慶行 (1995b). 意識調査の調査方式と調査結果その②, *ESTRELA*, 1995 年 11 月号, 34-42, (財)統計情報研究開発センター.
- 杉山明子 (1984). 『社会調査の基本』, 朝倉書店, 東京.
- 鈴木達三 (1996). 国際比較調査における標本計画と調査実施に関する一考察, 行動計量学, **23**, 46-62.
- 鈴木達三・高橋宏一 (1998). 『標本調査法』, 朝倉書店, 東京.
- 統計数理研究所国民性調査委員会 (1975). 『第 3 日本人の国民性』, 至誠堂, 東京.
- 統計数理研究所国民性調査委員会 (1982). 『第 4 日本人の国民性』, 出光書店, 東京.
- 統計数理研究所国民性調査委員会 (1992). 『第 5 日本人の国民性 — 戦後昭和期総集』, 出光書店, 東京.

On Sampling Scheme and Sampling Precision in Five Most Recent Surveys of the Japanese National Character Study

Tadahiko Maeda and Takashi Nakamura

(The Institute of Statistical Mathematics)

This paper explains the sampling scheme of the five most recent nationwide surveys (from the sixth to the tenth survey) of the Japanese National Character Study and presents the sampling precision of these surveys from two viewpoints: sex and age composition of the samples, and the magnitude of sampling errors. First, the sex and age compositions of both intended samples and achieved samples are compared with those of the estimated Japanese population in each survey year. In order to estimate the magnitude of bias caused by the deviation of the samples, achieved samples were post-stratified by sex and age and back-weighted by the stratum weight in the population to obtain a weighted estimate of the population proportion of the categories of survey items under consideration. Differences between unweighted estimates and weighted estimates by post-stratified samples are examined. The major findings are as follows: (1) Deviation from the population is larger in the achieved samples than in the intended samples. It grows larger in recent surveys. (2) Differences between unweighted estimates and weighted estimates are less than 1% for most of the categories. The differences are relatively larger in the tenth survey. Items with large bias in one survey also tend to show large bias in other surveys. Second, an estimate of the population proportion and its standard error are calculated, assuming that the sampling scheme in each survey was the stratified two-stage random sampling. Major findings are as follows: (3) The average value of standard errors is largest in the tenth survey because it was based upon the smallest sample size. (4) While the standard errors of most items in the sixth and the seventh surveys range between 1.1 and 1.2 times the standard errors assuming simple random sampling, the ratio of these two standard errors ranges from 1.2 to 1.3 in the eighth and the ninth survey. The results of the tenth survey are close to the latter group. (5) The average difference between weighted estimates of proportion and unweighted ones tends to be larger as the response rate becomes lower.

Key words: The Japanese National Character Survey, stratified two-stage random sampling, sampling error, variance of an estimator, ratio of standard errors, bias.