

国際比較における「データの安定性」に関する 一考察

— 中国調査データの検討を通じた文化多様体解析の試行 —

爨岩 晶¹・吉野 諒三^{2,3}・鄭 躍軍⁴

(受付 2007年1月10日; 改訂 2007年8月1日)

要 旨

本論文の目的は、中国 2002 年度調査(2002-2005 年度に遂行した「東アジア価値観国際比較調査」及び 2005 年度より進行中である「環太平洋価値観調査」の一部)において直面した標本抽出の問題点を検討することにより、意識の国際比較における回答データの安定性と調査文化の差異について、一考察を示すことである。この中で、意識の国際比較における方法論としての「文化多様体解析(CULMAN)」の一側面に触れる。

国際比較においては、調査対象となる各国・各地域における調査文化の違いに伴い、標本抽出方法の差異やウェイト補正に関する議論など、さまざまな問題が存在し、その中でどのように「国際比較」を可能にするのかが問われてくる。中国 2002 年度調査では、北京と香港において、日本と中国の調査文化の違いから、標本抽出計画と現実の調査での食い違いが生じた。この問題に対して、実際の回収データと標本抽出計画に近づけたデータ(加工データと称す)、センサスから得られた母集団のデータを比較するとともに、数量化3類を用いた国際比較を通して、データの安定性がいかに保証されるのかを、「文化の多様体解析」の立場から示し、将来の研究にむけて、特に我国における調査文化の変容を検討する際の手がかりを示唆する。

キーワード：中国価値観調査、文化多様体解析(CULMAN)、東アジア価値観国際比較調査、国民性、国際比較における統計的標本抽出、データの科学。

1. 「文化の科学」と「調査の文化」

各国の人々の意識を国際比較することは、各々の社会や文化を科学的に探る手段の一つであるが、さまざまな問題に出会うことがある。本論文の目的は、中国 2002 年度調査(2002-2005 年度に遂行した「東アジア価値観国際比較調査」及び 2005 年度より進行中である「環太平洋価値観調査」の一部)において直面した標本抽出の問題点を検討することにより、「意識の国際比較」における回答データの安定性と「調査の文化」(ある社会の文化における科学的調査に対する見方や態度を称して、ここでは「調査の文化」と呼ぶことにする)について、一考察を示

¹ 文部科学省 科学技術政策研究所：〒100-0005 東京都千代田区丸の内 2-5-1

² 統計数理研究所：〒106-8569 東京都港区南麻布 4-6-7

³ 総合研究大学院大学 先端科学研究科生命体科学専攻：〒240-0193 神奈川県三浦郡葉山町(湘南国際村)

⁴ 総合地球環境学研究所：〒603-8047 京都市北区上賀茂本山 457 番地 4

すことである。

まず、本研究の歴史的背景、その意義と展開について簡単に述べたい。

統計数理研究所では、1953年以來、半世紀以上にわたり、5年毎に成人男女を対象に「日本人の国民性」に関する調査を続けている。これは戦後民主主義の統計的基盤(世論調査の確立)と密接に結びついた研究でもあった(吉野, 2003, 2006a)。そしてこの研究は、1971年頃から、国民性をより深く考察する目的で、日本以外に住む日本人・日系人をはじめ、他の国々の人々との比較調査へと拡張され、現在遂行されている「意識の国際比較」研究へとつながっている。

「文化の科学」、特にその重要な方法論の一つである文化の比較を行う際、そこには「言語の差異」(例えば翻訳の問題は、距離的隔たりだけでなく、時間的隔たりにおいても無視できないものである)や、各国・各時代・各文化固有の調査方法の差異(その根底には、調査データや文献が依拠するパラダイムが存在する)といった「調査文化の差異」が存在し、そもそも異なる文化を計量比較することが可能なのかが問われてくる。特に意識調査では、いきなり全く異なる文化をもつ社会を比較しようとしても、計量的に深い意味のある結論を導き出すことはできない。

そこで、文化、言語や民族の系統など、何らかの共通点がある国々や社会を比較し、類似点、非類似点を判明させ、その程度を測ることによって、計量的な比較を行い、この比較の連鎖を拡張し、徐々に環として繋ぐことによって、やがてはグローバルな比較を可能にするという思想が生まれた。この思想の下で「意識の国際比較」研究が進められ、「連鎖的調査分析(Cultural Linkage Analysis, CLA)」と呼ぶ方法論(1978年米国調査において林知己夫が着想)が生まれたのである。そして、さらに比較を連鎖的につなげるだけでなく、様々な文化を、それを規定する様々な文脈の結節点として考え、多層構造を持ったものとして捉える「文化の多様体解析(CULMAN, Cultural Manifold Analysis)」(Yoshino, 2002; 吉野, 2005a)が目指されている。「意識の国際比較」を行おうとする者は、単純に調査回答データの表面上の数値のみを比べ、解釈するだけではなく、このような「国際比較可能性」を追求するための方法論を探求していかなければならない。「文化の科学」としての計量的文明論(林, 2000; 吉野, 2001)を確立するために、本調査グループでは「データの科学」(林, 2001; 林・山岡, 2001; 吉野, 2001; Yoshino and Hayashi, 2002)と称する統計哲学の下、様々な試行錯誤を行っているが、本論文もその一つである。

この20~30年ほど、国内外の機関により国際比較調査が数多く遂行されるようになったが、たとえ資金が十分であっても、統計的に厳密な標本調査がすぐに可能になるわけではない。政治的理由、各国の国内事情により、調査方法や調査の実施そのものが制約されるのである。つまり、文化を調査するには、調査自体(言語や調査方法等)を対象となる文化に合わせなければならず、そういった状況は無視して調査を遂行することはできないのである。例えば、1970年頃、統計数理研究所が最初に企画した国際比較調査であるブラジル日系人調査は、当時、軍政下のブラジル政府からビザが発給されず、急遽、ハワイ日系人調査へと変更されたエピソードがあった。また、世界の国々の中には、そもそも自国の正確な国勢調査データがなかったり、全国レベルの正確な戸籍簿や住民票などが一般には入手可能でなかったり、偏りのない国民を代表する適切な調査データを得ることが困難なところも少なくないのである。

本研究グループが今日までに調査した主な地域や国々を、表1に示す。

例えば、2002年度より4カ年、文部科学省・学術研究費補助金により、日本、中国(北京・上海・香港)、韓国、台湾、シンガポールの「東アジア価値観国際比較調査—「信頼感」の統計科学的解析」(East Asia Value Survey)を遂行した(吉野, 2004c, 2004d, 2006b)。2004年度からは、東アジア調査の第2ラウンドが統計数理研究所の基幹研究として開始され、さらにそれと関連させて、2005年度より4カ年の日本学術振興会・科学研究費補助金による「環太平洋価値観国際比較調

表 1. 統計数理研究所の国際比較調査.

実施年	各国での標本調査
1971	ハワイ在住の日系人
1978	ハワイ住民 (日系人を含む), 米国本土のアメリカ人
1983	ハワイ住民 (日系人を含む)
1987	イギリスのイギリス人, (旧) 西ドイツのドイツ人 フランスのフランス人
1988	ハワイ住民 (日系人を含む), 米国本土のアメリカ人 日本人 (A 調査)
1992	イタリアのイタリア人, ブラジルの日系人
1993	オランダのオランダ人
1998-1999	米国本土 (西海岸) の日系人
1999-2000	ハワイ住民 (日系人を含む)
2001-2002	中国 (北京, 上海)
2002	東アジア価値観国際比較調査 (East Asia Value Survey) 日本
2002-2003	中国 (北京, 上海, 香港)
2003	台湾, 韓国
2004	シンガポール
2005	環太平洋 (アジア太平洋) 価値観国際比較調査 (Asia-Pacific Value Survey) 北京, 上海, 香港
2006	米国本土, 韓国, 台湾 (遂行中)

注) すべて基本的には、全国レベルの無作為標本抽出法による面接調査である。ただし、ハワイ調査はホノルル市のみ、中国も北京や上海の中心市街地のみ、ブラジル調査はサンパウロを中心とする地域。上記の 1993 年までの調査については、「国民性七ヶ国比較」(林他, 1998, 出光書店)を参照せよ。

査」として、米国、オーストラリアやインドを含むアジア・太平洋地域の国際比較調査研究が進行中である。(既存の研究成果は一連の統計数理研究所・研究レポート等として発刊され、また、統計数理研究所の Web ページにある「研究紹介」(<http://www.ism.ac.jp/~yoshino/index.htm>)でも、最近の調査結果の一部が閲覧可能である。上記の時系列的かつ国際比較的調査の歴史と意義、「文化多様体解析(CULMAN)」の詳細については、Yoshino and Hayashi, 2002 や吉野, 2005a, 2005b を参照せよ。)本論文で扱うのは、「東アジア価値観国際比較調査」の一環として 2002 年度に行われた北京調査(以下、北京 2002 調査)と香港調査(以下、香港 2002 調査)である。

以下、第 2 節では、国際比較における調査文化の問題として、特に各国・各地域の標本抽出方法の差異について概説する。第 3 節では、回収標本に対するウェイト補正に関して、その一般的な問題点を取り上げる。第 4 節では、北京 2002 調査と香港 2002 調査における標本抽出計画と現実の調査での食い違いについて説明する。第 5 節では、この問題の検討のために、回収データとそれを加工したデータとの比較、それらとセンサスデータとの比較を試み、さらに回収データと加工データを国際比較の中で分析し、データの安定性について試行錯誤を展開させ

る。第6節は、それらの結果を「文化多様体解析」の立場から総括するとともに、第7節では、将来の研究にむけて、特に我国における「調査文化の変容」について触れよう。

2. 国際比較と各国の標本抽出法の差異の問題

先に述べたように、「文化の科学」の方法論の一つとして、異なる文化の社会の比較を可能にするには、まず調査文化の差異をどう扱うかが問われてくる。国際比較調査においては、翻訳の問題と各国固有の標本抽出方法の違い、それらが回答データに及ぼす効果に対していかに対処するかが基本的な課題となっている。

翻訳の問題へのアプローチとしては、翻訳・再翻訳(バックトランスレーション)の各段階でそれぞれ平行して複数のパイリンガルを活用した技法を試行しており、これについては既に論文や書籍で研究結果を発表している(林, 2001; 吉野, 2001; 吉野 他, 1995; Yoshino and Hayashi, 2002)。本論文では「調査文化の差異」に関する二つ目の問題、各国・各地域における標本抽出法の違いとそれが生み出す影響、そこでのデータの安定性について議論する。

世論調査のように「一人一票」の民主主義社会を体現するような標本抽出調査を行う際、統計的には調査対象となる諸個人が母集団を偏らずに代表するよう、無作為に選ばなければならない。例えば、日本人の成人全体を母集団として、3,000人の代表標本を抽出する際、各個人の全体から抽出される確率が同じになるように標本抽出計画を行う。日本では住民基本台帳や選挙人名簿が整備されているため、これらを利用した等確率の標本抽出方法が確立している。よく用いられる層別多段抽出法では、国勢調査データ等を活用して、成人の人口数や都市規模を考慮して、日本全体から調査地点を例えば150地点抽出して、各地点で、住民基本台帳や選挙人名簿から、統計的乱数を発生させて20名ずつ抽出し、全国で合計3,000名の計画標本を確定するのである。こういった方法を、統計理論的に標本抽出誤差(回収標本の統計量と真の値との差)を推定することができるという意味で、「科学的世論調査」と称している。

戦後の民主主義の発展のための科学的世論調査、その方法論の確立において、昭和20年代の初めから統計数理研究所を中心とした、マスコミ各社を含め、官民の世論調査機関の果たした役割は大きい。その歴史と実践については、今井(1996a, 1996b, 1997)、吉野(2003, 2004b, 2005a)などが参考になる。

しかし、日本以外の国や地域では、標本を抽出するための住民基本台帳や選挙人名簿などが存在しなかったり、存在したとしても政府の特殊な管理下におかれ、世論調査等には活用できなかったりすることがある。

欧州では、ランダム・ルート法と呼ばれる方法がよく用いられている。例えば本研究グループによる1992年のイタリア調査では、調査地点は国勢調査データに基づき確率比例抽出するが、各地点では、伝統的に(ローマ時代に各道路建設貢献者の名をつける慣習の名残か)小さな道を含めほとんどすべての道路に名前がついていることを利用して、道路のリストを地点のリストと見て、統計的乱数を発生させて特定の道路を抽出し、ランダム・スタート点を決め、その道路の例えば左側に沿って3軒おきに戸別訪問し、各戸では家族(成人)の中から誕生日法などで個人を特定して、面接調査を遂行する。拒否された場合は、次の3軒おいた家を訪ねる。これを例えば各地点で20名という目標数に到達するまで繰り返すのである。したがって、拒否を含む調査不能は記録されておらず、日本のように、計画標本数に対してどれだけ有効回答調査票が得られたかという、有効回収率は通常は計算されない。1988~1993年の日欧米7ヶ国国際比較調査(林 他, 1998)では、計画標本数の代わりに全アクセス数を分母として擬似的に回収率を計算したが、概略、欧州では30%強程度の回収率、したがって3人に1人は回答し、2人は調査不能(不在や病気などを含む)ということになる。

米国では、ミシガン大学などの幾つかの調査を別にすると、クォータ法(割り当て法)と呼ばれる調査が多く用いられている。この方法の場合、地点抽出は国勢調査データに基づき確率比例抽出するが、各地点では、あらかじめ指定された属性(性別、年齢層別)などに関しては国勢調査データに整合させて、各地点の標本総数 20 人中、例えば男の 20 歳台は 2 人、30 歳台は 3 人…、女の 20 歳台は 2 人、30 歳台は 4 人… 等と割り当て、該当する人間を探し出し、面接をするのである。問題なのは、実際の探し出し方にどの程度の恣意性や偏りが入るのか、ということである。ランダム・ウォークである道に沿って 3 軒ごとに訪問し、該当者を探す場合もあれば、常時面接回答者を集めやすいように、普段から、教会など集会所に頻繁に出席しておいて、顔見知りの人に依頼するようなこともありえる(あるフランスの調査会社でこの方法が用いられていた)。確かに、あらかじめ指定された属性(性、年齢、学歴、人種など)については、国勢調査データの属性分布から偏らない抽出方法ではあるが、しかし、指定されていない属性の偏りを防ぐことには何ら配慮がなされていない。属性は無限に考えられるし、特に当該調査で何が回答に大きな影響を及ぼす属性であるかは、あらかじめ分からないことが多い。そのような状況の中で、クォータ法は行われているのである。

また、クォータ法でも、ランダム・ルート法でも、通常、若年男性層の回答者が捉え難いのだが、国勢調査データの属性分布に合うよう、不足分のデータを補うために、回収データに「補正ウェイト」をかけ、事後に見かけ上、合理化するということが行われている。これについては、もともと偏ったデータ(回収された若年男性層は少ないだけでなく、回収・非回収層を含む若年男性層全体から偏った人々の可能性がある)であるのに、さらに想定外の方向の偏りを助長させる可能性があり、決して望ましくないと、再三、論じてきた(吉野, 2002, 2006a)。(これについては、以下でも触れる。)

いずれにせよ、ランダム・ルート法やクォータ法が用いられる調査文化の中では、標本抽出誤差という概念は成り立たず、したがって推定誤差は計算できず、統計理論的にはあまり望ましい状況とはいえないのである。

一人一票の選挙から成り立つ民主主義社会において、世論調査は歴史上、選挙以外の方法で世論や民意を知るために行われてきた。そういった意味で、世論調査では人々の意見を偏り無く調べるため、標本の等確率抽出(偏らない回答者の選び方)が理想とされるのであるが、この理想に近い標本抽出が可能であるのは、整備された住民票や選挙人名簿が活用できる日本の調査文化においてのみである(世論調査と民主主義の関係については、ESOMAR/WAPOR, 2007 や高橋, 2004 を参照せよ)。戦後、長年にわたり「日本の民主主義は民主主義ではない」という欧米からの批判(日本異質論)があったが、奇妙なことに、米国を含め、他の国々では、クォータ法など、統計学上は望ましくない標本抽出方法に甘んじているのが実情である。米国の 2000 年大統領選挙や 2002 年中間選挙における、マスメディアの世論調査に基づく予測が外れたのは、このことを象徴している。実質的な民主化は別としても、世論調査に限って言えば、日本が最も科学的であり、それゆえに民主主義的方法を使っていることになるのである。

ただし、日本以外では統計学的には望ましくない方法を用いているとはいえ、それは各国の歴史や伝統の中でそれぞれの国や社会で最善と思われる方法を工夫してきた結果である。例えば、統計的標本抽出理論が既に確立していた時代に、戦後の民主化とともに世論調査方法の発展をみた日本とは異なり、欧米の民主主義発展の歴史では、まず階級間、人種間、男女間の平等が謳われ、そのような社会では、アメリカの陪審制度の運営などに見られるように、利益の異なる集団間の平等に極めて敏感だったのであり、集団間の権力の適正な「割り当て(クォータ)」が求められていた。その大義の前では、同じ階級や人種、性別の内部でも個人差が存在するという現実、小さなものとみなされるであろう。それぞれの国や地域は、そのような異なる調査文化の中で得られた調査データに基づいて、政策立案や政策評価に繋がる判断をして

いるのである。したがって、狭量な統計理論の見地からのみ非難するのは、むしろそれ自体が狭量な見方であり、各国・各地域の調査文化を尊重していくべきであろう。

上記のような文化と歴史の差異に基づく調査文化の差異は、意識の国際比較を行う際、各国・各地域で異なる標本抽出法によって得られた回答データをどのように解析していくべきであるのかという問題に関連してくる。

ここで、本節の最初にあげた、林 他(1998, p. 17), 林(2001, pp. 74-81), 吉野(2001), Yoshino and Hayashi (2002)で述べられている国際比較調査におけるデータ解析から得られた成果、特に翻訳の問題や、実際の回収データと無作為標本抽出との比較を通じた分析(この場合の比較は、「調査文化」間ではなく、特定の「調査文化」内で異なる標本抽出方法を用いて得たデータの比較である)、その際のデータの安定性について要点を記しておこう。(本論文では、「調査文化の差異」について、翻訳の問題ではなく、標本抽出の問題のみを扱うのだが、方法論的には通ずるものがある。)

1) 単純集計表での一つ一つの質問項目毎の回答分布の比較では、標本抽出法の差異や、翻訳過程の微妙な表現の差異で回答結果に10%から15%程度の差が生じることもあり、その程度の差異では直ちには本質的な意味のある差異とは結論できないこと。

しかしながら、

2) 複数の質問項目群に対応する複数の国々の回答データのパターン解析(数量化3類や多次元データ解析)による国際比較では、標本抽出法の差異、事前抽出確率また事後の回収標本の偏りに関するウェイト補正の有無、質問項目の表現の多少の差異、一部の項目の入れ替えなどに対しても、それらの影響は無視できる程度のものとなり、かなり安定した結果が得られることも多い。

したがって、

3) 単純集計表を用いた比較でも、一度、複数の項目群のクロス集計や多次元データ解析などでデータの安定性を確認してから、当該の単純集計表やクロス集計表に戻り、差の有無を論ずる慎重さが必要である。

上記の1)から3)を念頭に、以下の標本抽出法についての問題とその解決策の試行を考えよう。また海外調査データの安定性(回収データの事後のウェイト補正の有無の問題)については、鄭(2003a, 2003b)や山岡・李(2004)も参照せよ。

3. 有効回収標本の属性分布の母集団からの偏りについて

ここで、次節以降で行う試論をわかりやすくするため、標本抽出とウェイト補正に関する一般的な問題点を概説したい。

通常、日本の世論調査などで行われる住民基本台帳や選挙人名簿を利用した標本抽出では、最終抽出単位となる各個人は、母集団の中から等確率で抽出される。この点が、回収標本全体の平均値をもって母集団(例えば日本人の有権者全体)の真の値の推定値とする根拠となる。しかし現実には、不在、拒否、病気や死亡、移転等々の理由から、計画標本の全員から回答が得られるわけではない。つまり、有効回収率が100%を前提として標本の統計量を母集団の真の値の推定値とするのは当然だが、現実には有効回収率が100%ということはまずありえず、これは統計学に基づく標本抽出理論のフィクションに過ぎない。ただし、有効回収率が100%からあまり逸脱しない限りで、このフィクションは十分に有効であることが確認され、長年用いられてきた。例えば、戦後日本の民主主義を発展させるための科学的世論調査研究の歴史とも関連する、1953年の「日本人の読み書き能力調査」の準備の中で、神奈川県小田原市での全数

調査の有効回収票全体（母集団）の値と、そこから無作為抽出した回収票（標本）に基づく推定値とを比較して、標本抽出理論の有効性が証明されている（今井, 1996a, 1996b, 1997; 吉野, 2003）。

数年前までは政府の世論調査でも 70% 回収を目標に調査計画が立てられることが多かった。しかし、年月を経るにつれて有効回収率は低下し続け、さらに 2005 年の民間調査会社の不祥事で、厳密に有効回収率を計算してみると、特別な工夫がなされていない通常の政府の調査では 50% 台に落ちていることが判明した。

この有効回収率が 100% に遠く達しないデータでも、計画標本全体の中で、非回収層と回収層の意見の分布があまり異ならないのであれば、問題は小さい。しかし、そもそも、非回収層の意見は観測不能なのであるから、その差を直接確かめることは原理的にできない。また、特別なケースは別にして、一般には常に差が小さいと想定するのは妥当とは思えない。（この議論については、吉野, 2006a を参考にしていきたい。）

3.1 母集団の属性に基づくウェイト補正

このような低い回収率の調査ですぐに指摘されるのが、性別や年齢の偏りを「補正」として、性別や年齢層の分布を計画標本や国勢調査の分布（母集団）に合うように回収データにウェイトをかけ（前者では全体として男性のほうが少なく、特に若年層の男性が少なく、主婦が多くなる傾向がある）、各質問項目の回答分布にも同じウェイトをかけることである。しかしこれでは、そもそも偏りのある回収層の意見に、さらに様々な潜在的要因を引き込み、予想もつかない複雑な偏りを加えてしまう可能性がある。例えば、通常、単身居住で不在も多く、回答者として捉えにくい男性の若者の場合、たまたま病気や失業などで在宅していた人の回答がウェイトをかけられ、男性若年層の代表的意見とされてしまうこともありうる。これが適切でないことは、すぐに分かるであろう。もとの回収データのままだよというわけではないが、偏りがある元々のデータ以上に、偏りがどちらの方向へどれだけあるかが推定できないようなデータの方が扱い難いのである（吉野, 2002）。

3.2 標本抽出計画に基づくウェイト補正

他方で、住民基本台帳などの名簿を用いないエリアサンプリング（住宅地図から世帯を抽出）や RDD 電話調査などでは、別の意味でのウェイト補正が問題となる。これらの方法では、（少なくとも計画上は）世帯抽出における各世帯の抽出確率が等しく、さらに各世帯の成人の中から誕生日法などで個人を抽出する際、その家族成員の中では等確率に抽出されることが意図されており、その結果、最終段階での各個人の抽出確率は、各家庭の成人数に反比例することになる。それゆえ、少なくとも計画段階では、この方法で抽出した個人データを、対応する世帯内の成人数に比例したウェイトをかけて補正することが求められる。このウェイト補正は、先の回収標本を母集団の属性分布に合わせる見かけ上の補正とは異なり、理論に基づく当然の要請である（その意味では、「補正」ではなく「ウェイト付け」と呼ぶべきかもしれない）。

しかし、有効回収率が 100% のときは問題ないが、現実の調査では 100% からかなり下回ることが多く（やはり比較的男性の若年層が少なく主婦が多い）、実際の有効回収データに対して、世帯内の成人数に比例したウェイト補正を施すと、単身居住男性のウェイトは小さく、主婦のウェイトが高くなり、もとの偏りをさらに増長する危惧がある。

エリアサンプリングにおいては、十分に有効回収率が高ければ、大きな問題はないのかもしれない（その場合、計画標本に対する非回収数をデータに明示しなければならない）。しかし、今日行われている RDD 電話調査のように（実質の回収率はかなり低く）、あらかじめ決められた目標回答者数に達するまで、その間に幾人に拒否されても調査を続けるという方法では、この種のウェイト補正を施すと、さらに偏りを呼び込む危惧がある。ある新聞社の RDD 電話調

査では、先にあげた事後の回収データのウェイト補正と、ここで述べた世帯ウェイト補正の両方を用いているが、両ウェイト補正によって想定外の偏りが起きている可能性がある。

ヨーロッパ世論・市場調査協会、世界世論調査学会のガイドライン(ESOMAR/WAPOR, 2007, p. 23)では、「正確なサンプル・バランスを得るために、簡単な人口統計に基づくウェイト補正を行うことは、通常良い方法である」としている。これに対し、本田・本川(2005, p. 233)は、回収率の低い層(例えば若い男性)を補うウェイト付けは、その層のウェイトが大きくなり、誤差が非常に大きくなる、ということを指摘している。Kish(1992)は、目的によってはウェイト補正が望ましい場合とそうでない場合があり、妥協点を見出すためには更なる理論が必要であると述べている。

いずれにせよ、元の回収データのままでよいというのではないが、本稿で取り扱う中国2002調査をはじめ、本調査グループが遂行している「意識の国際比較」における調査データは、母集団属性に基づくウェイト補正はせず、標本抽出計画に基づくウェイト補正も、回収率が低い場合、偏った意見を強調してしまう可能性を考慮して行わず、そのままのデータを解析に用いている。

4. 北京調査と香港調査の標本抽出計画と問題点

さて、本節と次節では、統計数理研究所が主体となって実施した「東アジア価値観国際比較調査」における北京調査と香港調査で起きた事例をもとに、標本抽出の問題点とデータの安定性について具体的に検討する。

理想的には、各国・各地域で厳密な無作為抽出法(回答者個人レベルの等確率抽出)によるデータ収集が望ましいことは勿論だが、費用や時間の制限などの調査自身の限界、各国・各地域での調査文化による制限を勘案すると、現実的に不可能な無作為抽出法にこだわるよりも、普段現地で用いられている方法を用いた方が無難という判断もあり得る。実は、この点の重要性が再認識される事態となったのである。

まず本節では、北京調査と香港調査における標本抽出計画とその抽出結果の問題点(本調査グループと現地調査機関の担当者とのコミュニケーションの錯誤で、当初計画していた等確率抽出に回収データがなっていない可能性がある)について説明する。そして次節では、試行的に元の回収データといくつかの方法で修正して本来計画した無作為抽出に近づけたもの(厳密な意味では無作為とは言いがたいが)とを比較検討し、どの程度の差が存在するのかを分析するとともに、数量化理論に基づく多次元パターン分析を通して、データの信頼性や安定性について試行的な議論を行う(分析については、Yoshino, 2006を参照)。

4.1 北京調査

中国では、政治経済の目覚ましい発展とともに、社会の様々な側面が著しく変化している途上でもあり、これは社会調査の環境についても同様である。将来のことはともかく、以下では、調査時点において種々の条件を勘案し、ベストではないが限定された条件の下で採用した標本抽出の「操作的手続き」である。

中国本土における調査では、国家体制が日本とは異なっており、調査地域において住民基本台帳や選挙人名簿が一般には非公開のため、それらに基づく標本抽出は望めなかった。そのため、この調査に先立って実施された北京・上海市民意識調査(鄭, 2003a, 2003b)の経験を踏まえ、北京市と上海市とが管轄する都市中心部の区に限って(各市とも全域ではない)、一種のエリアサンプリングの形で、第1次抽出単位として日本の自治会に相当する居民委員会・村民委員会(以下、居民委員会と通称する)を50地点分抽出し、第2次抽出単位として各居民委員会から20世帯を抽出し、第3次抽出単位として各世帯から個人を1人抽出するといった、3段階抽出法

に基づく標本抽出計画を立案し、そこで抽出された個人に対して面接調査を遂行した(2002年11月~12月)。以下、その詳細である。

1) 第1次抽出

まず、調査地域を構成する居民委員会のリストから、人口規模に比例して調査地点となる居民委員会の抽出作業を行う。その際、調査地域の居民委員会の異動や調査不能に備え、最初に計画した地点数の2~4倍を抽出しておき、そこから計画調査地点を第1次抽出単位として選び、残りは予備調査地点とする。実際の調査段階において、何らかの理由(注. 北京オリンピックの準備に関連する住居の強制大移動や新住居の建設などをはじめ、行政的にも地理的にも中国大都市部は急変動している)で調査不能となった地点があった場合、その地点に最も近い予備調査地点で代替させ、計画調査地点に組み入れる。

本調査では、第一次抽出単位の計画標本として抽出した50の居民委員会のうち、開始時点または途中でビル管理者による干渉で調査を中断したところが7箇所あり、それを代替した7箇所の居民委員会を加えて、実際に取り扱った居民委員会は57であった。

2) 第2次抽出

次に、第2次抽出の準備として、選ばれた居民委員会を調査監督者(複数)が訪ね、各々の居民委員会が管轄する総世帯数を確認するとともに、調査地点の住居数や交通でのアクセス方法なども明記した住宅配置図(例は図1を参照)を作成する。この総世帯数の情報と住宅配置図を用いて、第2次抽出単位としての世帯を調査地点から抽出する。その具体的方法は、はじめに、計画調査地点となった全居民委員会の総世帯数を計画標本の数で割り、そこで求められる数を訪問間隔(北京では36軒ごと)とする。そして、世帯抽出員が各調査地点において、調査監督者によって無作為に決められた起算世帯住所(スタート点)から、先の訪問間隔で抽出される世帯に1,2,3,...のような番号をつけ、調査実施の進行順を示す矢印とともに住宅配置図上に明記し、訪問面接調査のための調査対象世帯リストを作成する(該当する居民委員会の全体を等間隔で偏らずにカバーするようにする)。

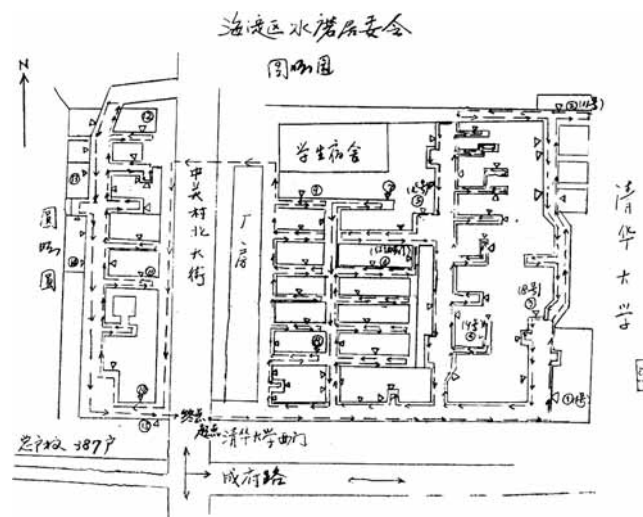


図1. 住宅配置図.

表 2. Kish-Grid 法を用いた標本抽出表. これを 3 種類用意して, 各居民委員会から抽出された各世帯で成人 1 名を抽出した.

隨機抽樣表 (A)

成員 編號	符合條件的家庭成員				地址編號 ()									
	稱謂	出生日期 年/月	性別	選擇	01	02	03	04	05	06	07	08	09	10
1					1	1	1	1	1	1	1	1	1	1
2					2	1	1	2	1	2	2	1	1	2
3					3	2	1	2	1	3	1	2	3	2
4					4	1	2	3	3	4	1	2	4	2
5					5	4	3	2	1	2	3	4	5	1
6					6	5	1	2	4	3	1	4	5	6
7					5	3	2	1	6	7	2	4	3	7
8					7	6	5	4	3	2	1	8	1	2
9					8	9	7	6	5	4	3	2	1	6
10					9	8	7	10	5	4	3	2	1	0

注意事項: 1. 家庭成員の登記序號按照年齡大小“從大到小”的順序依次排列.

2. 符合條件家庭成員應同時具備下列條件: 18 歲以上的民眾.

3. 選擇依據被抽中樣本的抽選序號的尾數與符合條件的家庭人口數交乘的數位而確定相應序號所對應的家庭成員, 並在該家庭成員同一行的“選擇”欄中打“√”.

(注 調査対象抽出用乱数表の利用手順: 1. 家族構成員の登記は年齢の高い方から低い方へ順番に記入する. 2. 条件に合う家族構成員とは以下の条件を全て満たすものである. ①成人 (18 歳以上) である. ②調査対象地域に戸籍を置いて 1 年以上居住している. ③毎週平均 5 日以上この家で暮らしている. 3. サンプルングは抽出された世帯の世帯番号下 1 桁の数字と条件に合う家族構成員の人数を考慮した数字によって乱数表から該当家族構成員に該当する番号を確定し, その家族構成員の行の“選択”欄に調査対象個人として“√”をつける.)

3) 第 3 次抽出

最後に, 抽出された世帯から乱数表を利用して個人 (第 3 次抽出単位) を無作為に抽出する. 調査対象世帯で直接, 面接調査を行う調査員は, 番号 (地址編號) の下 1 桁の数字と, 対応する世帯の成人家族構成員数 (生年月日の降順) をもとに, 表 2 のような調査対象者抽出乱数表を使って個人を特定し, 面接の調査対象にする (Kish 法).

たとえば表 2 では, 15 番の標本世帯に成人構成員が 4 人いる場合, 乱数表の第 4 行 (成人構成員数) と第 5 列 (世帯番号の下 1 桁の数字) が交差するセルの数字が 3 となっているため, 調査対象は 3 番目の構成員と決定される. なお, 各居民委員会において, 世帯番号の下 1 桁の数字 1, 2, 3, ..., 9, 0 は一様に分布していない可能性があり, 例えば 1, 2, 3, ..., 9, 0 の順で生起する確率が減少していくようなこともあるかもしれない. したがって, 調査対象抽出用の乱数表については, 世帯番号の分布と家族構成の特性を考慮した上で, 合わせて A, B, C の 3 種類の乱数表を作成し, 調査票全体に均等に割り当てるようにする.

ここで注意しておくが, 各国の都市部の現状では避けられない回答協力拒否や回答者の不在と, そこから生じる低回収率の問題に対して, 中国本土における調査では, 費用や調査日数などの様々なコストを勘案し, 地点によっては, やむを得ず代替標本を用いることを許している. ただし, 各計画標本 (ここでは世帯) に対して拒否, あるいは 3 度の訪問でも不在, 病気などの

表 3. 調査票回収状況(北京 2002 調査).

最終 使用 標本数	有効 回収 標本数	拒否				不在		接触 不可	非該当	廃屋 移転
		本調査 に興味 なし	多忙	全ての 調査拒 否	その他	長期	一時			
3634	1062	190	811	600	300	57	126	371	42	75

理由で協力が得られない場合に調査不能と判断して、対応する代替標本を、例えば隣の3件目の世帯から抽出し、それでも協力が得られない場合は、その手続きを繰り返すなど、調査員の恣意性を排除する代替標本の取り方をあらかじめ明確に指示しておいた。このため、厳密な無作為標本抽出にはなっておらず、どちらかといえば、「地点抽出は人口確率比例の無作為抽出で、各地点での個人抽出は、恣意性は排除しているが、あらかじめ定められた調査対象者数だけを割り当てて調査する」方法に近い抽出計画となっている。

なお、予備標本の導入分を含め、実際には3,634世帯を訪ねて調査を試みた。回収した有効数は1,062名であった。日本における無作為抽出法を模して、訪問世帯数を分母として計算した擬似的な回収率は29.2%に留まり、調査不能は69%にのぼった(回収状況については表3を参照)。参考までに述べると、これは欧州などにおけるランダム・ルート・サンプリングや割り当て法の場合と同程度である。

4.2 香港調査

香港における標本抽出計画は、香港城市大学の研究者を現地研究協力者とし、彼らとの検討を経て以下のように確定した。(同大学は学術的にも商業的にも各種の調査を遂行しているが、CATI [Computer Aided Telephone Interview] 法が主流で、今回は、我々との協議を経て、面接調査を遂行してもらった。)

香港城市大学では、各種の調査経験が豊富であり、特に香港統計局とも連携して家計調査を遂行してきた経験がある。そのような調査では、TPU (調査地区単位) は、各地域の経済的特性(高級高層住宅街か否かなど)を考慮して区分されている。1997年の英国からの返還による住民の移動(海外へ流出後、帰還)やビジネスのための短期的移動(香港と大陸内部との間)の多さを考慮して、香港、マカオ、台湾を含む中国に5年以上住んでいる人々のうち、現在、香港に居住する成人(18歳以上)を対象とした。

あらかじめ実際の有効回収率が予想できなかったため、TPU (調査地区単位) を考慮した層別無作為抽出により、2,000世帯と1,000世帯のセットを2本作り、はじめの2,000世帯のセットを用いて訪問調査するが、回収数が不十分な時は次の1,000世帯のセットを用いることと決めていた。実際には両方とも利用したため、結局、計画サンプルとして3,000世帯が抽出され、調査されたことと同じになった。全体の手続きは以下のとおりである。

1) 第1次抽出

香港の全282TPUから、各TPUの人口に比例して、50のTPUを確率抽出する(当地の統計局の協力による)。

2) 第2次抽出

抽出された50のTPU全体から人口比で3,000のLB (Living Block) を等確率抽出し、計画サンプルとして合計3,000世帯を抽出する。(LBは一戸に対応し、例外的に1LBに複数の世帯もあるが、概ね90%は、1LBが1世帯に対応すると報告を受けた。)

表 4. 調査票回収状況 (香港 2002 調査).

計画 標本数	有効回収 標本数	無効票	拒否	不在	中止	商業地	廃屋 移転	言語 問題	非該当
3000	1057	39	1336	325	21	15	165	30	12

3) 第3次抽出

抽出された各世帯から、その家族の中で18歳以上、かつ5年以上香港に居住するもの全員に対して Kish 法を用い、最終的に面接の対象者として家族から1名を決定する。(計画サンプルとして抽出された各世帯が留守の場合は、異なる時間や日を見計らって、最低5回は接触に努めた。)

予期せぬこととして、調査遂行(2002年12月~2003年3月)の末期において SARS 発生のため、調査を早く切り上げるという事態があったことには注意が必要である。特に、「病気」や「不安感」に関する調査項目などでは、データ解析上、その影響を考慮すべきである。

最終的な有効回収票数は1,057(回収率35.2%)である。詳細を表4に示す。

4.3 問題の所在

以上、北京と香港の標本抽出計画の概要を述べた。住民基本台帳や選挙人名簿等を見ることができず、世帯ごとの構成員数も把握できない以上、第3次抽出までを加味した「等確率抽出」は不可能であったが、第1次抽出と第2次抽出、つまり世帯の抽出までは「等確率」になるよう計画していた。中国本土の調査では、第1次抽出単位(地点)を世帯数に応じて確率比例抽出しているので、第2次抽出単位(世帯)では、全地点で「同数」を抽出すべきである。しかし、最終的に報告されてきた各地点での回収状況表を見ると、これが必ずしも一定とはなっていない(吉野, 2004d, pp. 8-17)。可能性として、現地の調査監督者が当初予定していた抽出計画を誤解し、第2次抽出単位(世帯)が「同数」にされなかったのではという懸念がある。特に、確率比例の考え方を勘違いして、各地点での抽出世帯数をも、全抽出地点の総世帯数に対して比例させたのかもしれない。これでは、比例の重み付けを2重にしてしまったことになり、「等確率抽出」にはならない。

本論文では説明を省いたが、同時期到北京と同じ方法で調査した上海では、回収率が低いため計画標本をすべて回収できず、ほとんどの地点で20前後の回収数で調査を打ち切り、結局は、第2次抽出が「同数」を擬した形になっている。(強引ではあるが、北京の方を取って正当化するならば、信頼できる直近の人口分布データがなかったために、第1次抽出単位である地点を「等確率抽出」し、第2次抽出単位である世帯を総世帯数に比例させた「確率比例抽出」で行っていたと捉えられないわけでもない。)

香港の場合も、各 TPU での計画標本サイズが異なって報告されているので、比例の重み付けを2重にしてしまった可能性がある。事後に、現地調査監督者に聴取した限りでは、手続き上で、抽出された50TPU全体の総世帯数に対して各 TPU の世帯数を比例させ、それに基づいて3,000LBを抽出した可能性が高い。協力してくれた現地統計局が第1次抽出単位である TPU を抽出し、情報提供してくれたようであるが、その TPU 抽出が「等確率抽出」になっていたのか否かの詳細は、今となっては不明である。

いずれにせよ、統計学的に厳格な方法に固執し、現地の「調査文化」にはない(ゆえに現地の調査機関が慣れていない)標本抽出方法の遂行を依頼したが、予想外の誤謬が入り込んだ危惧を捨てきれない。データ解析の際は、これを念頭において、ある程度の誤差や標本抽出法の

差異に対して安定した結果を得る方策を勘案しなければならない。

次節では、かなり限定された条件や仮定の下ではあるが、この回収データの偏りに関して考察を加える。

5. 回収データと加工データとの比較

ここでは、北京 2002 調査と香港 2002 調査における回収データと、それを加工して各地点での標本数をほぼ揃えたデータとの比較を行う。今回の調査結果に対しては、以下に示す三つの方法で、三種類の加工データを作成した。いずれの方法も、各地点から一定数の標本を取り出した状態を作ることで、第 2 次抽出で行われた可能性のある 2 重の重み付け(各地点内の人口や世帯数によって標本数を調整すること)から生まれる誤謬の影響を軽減する試みである。

もっとも前述のように、サンプリング状況に不明な点があるため、回収データと加工データのどちらが真の推定値に近いかわからないが、両者の差の大きさを見ることで、分析におけるデータの安定性を考察する手がかりとしたい。

① ウェイト補正データ(回収データの確率補正)

各地点のサンプルサイズの違いに対し、地点ごとに一定のウェイトをかけ、各地点での標本数の大小の影響を除くようにデータを確率補正した。具体的には、各地点のサンプルサイズがすべて 20 の場合と同様になるようにウェイトをかけた。第 3 節で紹介した「標本抽出計画に基づくウェイト補正」に近い形であるため、ここでは「ウェイト補正データ」と呼ぶ(ただし、本来の手続きが厳密な確率抽出になっていないことや、「母集団属性分布に基づくウェイト補正」とは異なることに注意)。

② 20 再割当データ(本来の標本抽出計画を考慮したデータの加工)

意図された元々の計画では、各地点から 20 人抽出することになっていた。そこで、本来の計画通りに、回収データから各地点 20 人分のデータを取り出した。当初の計画では、20 人分のデータを得た時点で終了することを考えていたため、それに合わせるよう、調査順の早いものから 20 人分のデータを取り出した。各地点に 20 人分のデータを割り当てるという意味で、「20 再割当データ」と呼ぶことにする。

このデータが一番当初に意図した計画に近いが、結果は北京 2002 調査の場合、サンプルサイズが 20 以上なのは 23 地点であり、残りの 30 地点は 20 に満たない。香港 2002 調査では、20 以上が 21 地点、20 未満なのが 25 地点となってしまった。これら、20 に満たない地点は、その全データをそのまま用いている。

③ 10 再割当データ(本来の標本抽出計画、及び回収結果を考慮したデータの加工)

20 再割当データと同様の方法で、各地点から 10 人分のデータを取り出した(よって「10 再割当データ」と呼ぶ)。北京では 53 地点中 16 地点、香港では 46 地点中 15 地点が 10 に満たない。20 再割当データに比べて、サンプルサイズがかなり小さくなるが、各地点間でのサンプルサイズの差が小さくなるので、比較の意味で試行する。

いずれも絶対的に適正な「補正」などではないが、あえていうと、①は回収データの「数学的補正」、②と③は回収の手続きが本来の標本抽出計画に近い形にした「データの加工」を試行した。

5.1 回収データと各加工データとの単純集計比較

北京 2002 調査と香港 2002 調査それぞれにおける回収データと、上述の方法で作成した加工データとの回答比率を比較しよう。まず、各質問項目の各回答カテゴリーの選択率に関して、単純集計表レベルで見る。図 2 は、元データの比率を横軸にとり、各加工データの比率を縦軸

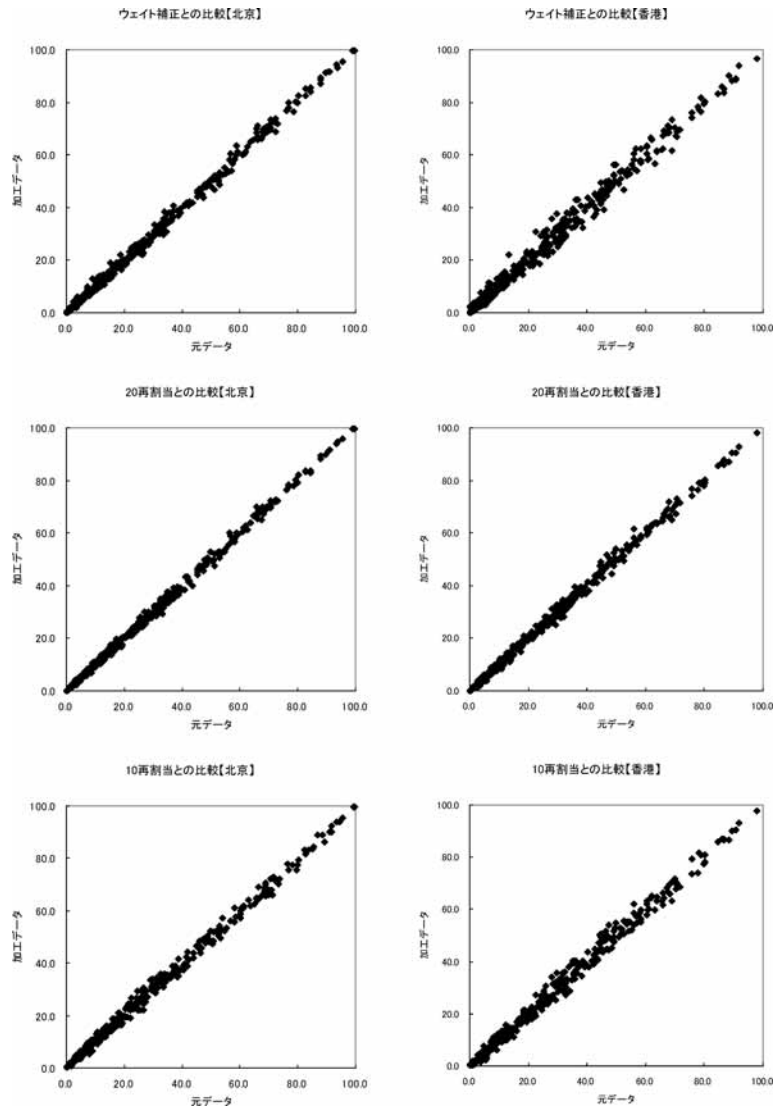


図2. 回収データと加工データの比較. 左が北京調査, 右が香港調査で, 上からウェイト補正データとの比較, 20再割当データとの比較, 一番下が10再割当データとの比較になっている.

にとって, 全回答カテゴリーを散布図にしたものである. 左が北京調査, 右が香港調査になっており, 上からウェイト補正データとの比較, 20再割当データとの比較, 一番下が10再割当データとの比較になっている. 両者が完全に一致すれば, 45度の直線上にのるのが, 概略としてこの直線付近に分布しているのがわかる.

元の回収データと各加工データの全回答カテゴリーにおける選択率の差の絶対値をとり, それらすべてを平均した値を比較してみる. 北京と香港, それぞれの回収データと各加工データとの差の絶対値の平均, その標準偏差, 最大値は, 表5の通りである.

表 5. 加工による差の比較.

	北京			香港		
	平均	標準偏差	最大値	平均	標準偏差	最大値
ウェイト補正	1.00%	0.97%	5.21%	1.86%	1.63%	8.73%
20 再割当	0.80%	0.74%	3.91%	0.95%	0.92%	5.70%
10 再割当	1.22%	1.06%	5.29%	1.54%	1.33%	6.40%

回収データとの差については、両調査ともに 20 割当データが一番小さいことがわかる。これは、先の図 2 からも見取れる。北京の場合は 10 割当データ、香港の場合はウェイト補正データにおける差が比較的大きい。また、全体としては、北京よりも香港の方が、差が大きいことがわかる。

ちなみに、この規模(元の標本数が 1,000 ぐらい)の「単純無作為標本」においては、回答比率 50%での標本抽出誤差は約 1.5%, 95%信頼区間は約±3.0%であり、さらに多段抽出であれば、その約 1.3~2.0 倍程度の標本抽出誤差が見込まれる。本調査では、先に述べたように、正確な標本抽出誤差は計算できないため、あくまでも参考にしかならないが、回収データと各加工データとの差は、その標本抽出誤差の範囲にほぼ収まっているといえる。(実際には、さらに非標本抽出誤差を標本抽出誤差とほぼ同程度に見積もることも多く(林, 1984), いずれにせよ、上記の差は許容誤差の範囲といえよう。)

5.2 各データとセンサスデータとの比較

各質問項目の回答分布において、回収データと加工データの差は、それほど大きいものではなかった。次に、属性項目(性別と年齢階層)を、これに対応する各地域のセンサスデータと比較してみる。なお、香港のセンサスデータは、調査機関と同じ 2002 年のものを使っているが、北京の場合は入手できたセンサスデータが 2004 年であり、2 年のズレがあることに注意していただきたい(しかもセンサスデータは北京全体であるが、本調査の対象地域は北京の都市部に限定されており、あくまでも参考としての比較である)。また、センサスデータの比較においては、年齢階層を見やすくするために、20 歳未満の者、70 歳以上の者を除外して各データを再集計している。

図 3 は、センサスデータ(図中は「census」)、回収データ(図中は「2002」)、加工データ(ウェイト補正は「weight」、20 再割当は「20」、10 再割当は「10」)における男女比を表したものである。

北京の回収データでは、センサスと比較して男性が約 6% 少なめになっており、各加工データでは若干その差が少なくなっている。香港のデータは、先述したとおり回収データと加工データの差が比較的大きかったが、センサスと比べてみると、北京ほどの違いが見られない(10 再割当での差が大きいのが、2%程度)。しかし、よし悪しは判断できないが、センサスとの違い、加工の効果の内実は、男女別に年齢階層の比率を比較してみるとより鮮明になってくる。まずは、北京のデータにおける年齢階層比とそれを男女別に分けた図 4 を見ていただきたい。(年齢階層の曲線を見やすくするために、70 歳以上を除外した。)

北京の回収・加工データは、センサスデータと比較して、単純に年齢階層だけで見た場合は 40 歳代以下の者が少なく、40 歳代後半と 60 歳代前半の者が多くなっているが、回収データ、加工データともに似たような傾向を示している。しかし、男女別で見た場合、男性では 30 歳

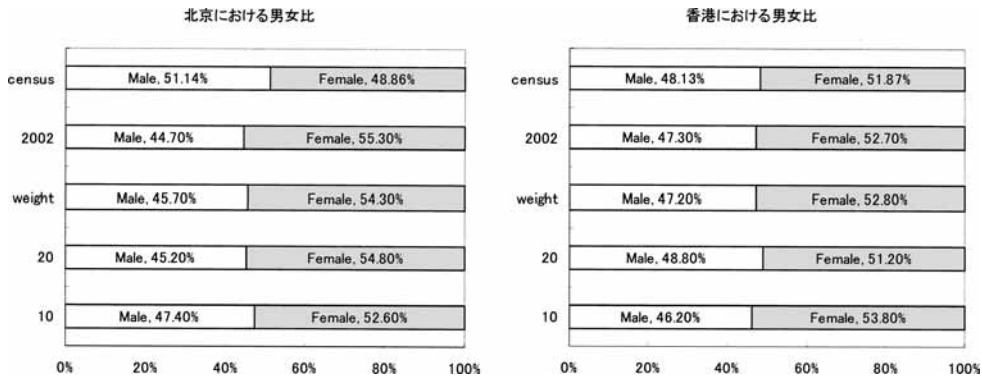


図3. センサスデータとの比較(男女比).

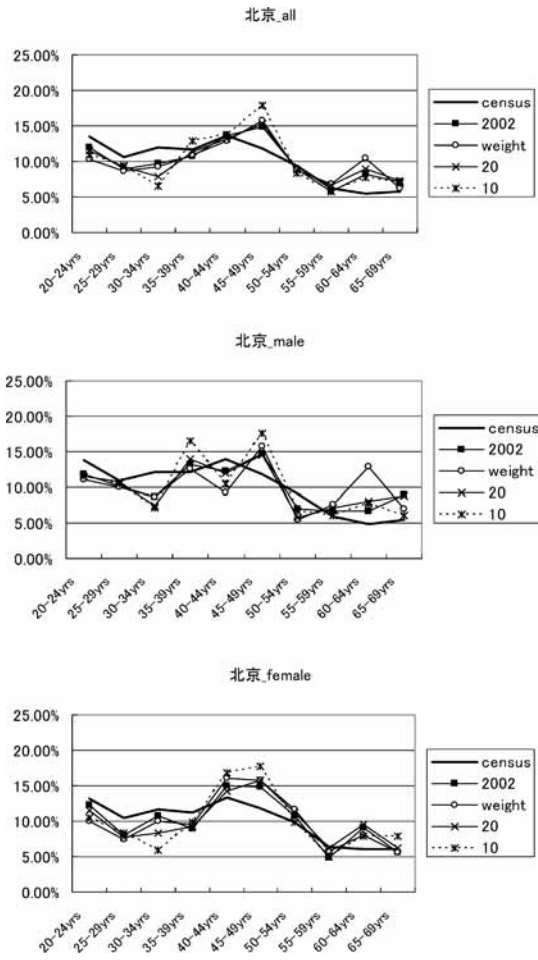


図4. 男女別年齢階層比の比較(北京).

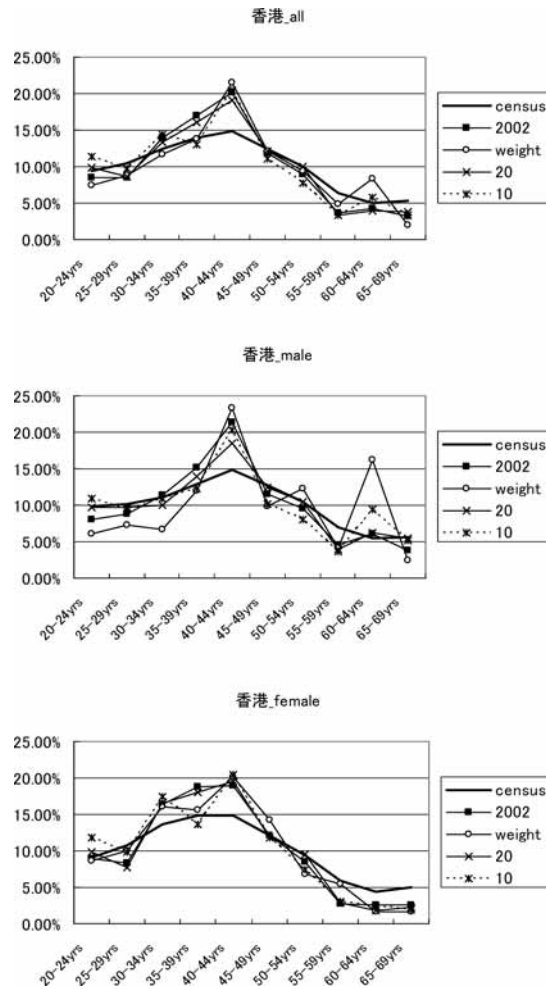


図 5. 男女別年齢階層比の比較 (香港).

代前半が少なく、40 歳代後半と 60 歳以上が多いという傾向は同じであるが、その違いは大きくなっており、このような傾向は女性の場合にも見られる。加工データに注目すると、20 再割当はそれほどでもないが、10 再割当データは回収データとセンサスデータとの違いを強調していることがわかる。ウェイト補正データにおいては、特に男性の 60 歳代前半のセンサスデータとの乖離が顕著である。つまり、単純に年齢階層を見たとき以上に、性別とクロスしてみた場合は、センサスデータとの違いも、回収データと加工データとの違いも広がっていることがわかる。

次に、香港データの場合を図 5 に示す。

年齢階層のみで見た場合、センサスと回収・加工データの間には、30 歳代後半から 40 歳代前半において大きな差が見られる。この傾向は、男女別で見た場合にも存在するが、特に男性の場合に注目すると、20 再割当データがセンサスに近づいているのに対して、ウェイト補正データでは、特に 60 歳代前半でセンサスや回収データとの顕著な差異が見受けられる。

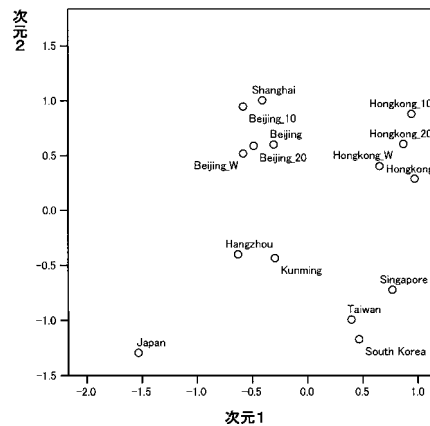


図 6. 性別、年齢階層における国際比較。性別、年齢階層、国・地域を数量化3類によって分析し、国・地域のみをプロットした。北京の回収データは「Beijing」、ウェイト補正データは「Beijing.W」、20再割当データは「Beijing.20」、10再割当データは「Beijing.10」とし、香港も同様に表記した。

以上から言えるのは、男女比のセンサスに対する差は少ないにしても、性別と年齢階層でクロス集計した場合のセンサスとの差はかなり大きい、ということである。調査データは、たとえセンサスであったとしても各国・各地域における「調査文化」の影響を受けており、それに一致するかどうかでデータの良し悪しを判断するのは短絡的と言われるかもしれない。ただし、「調査文化」に影響された元々の回収データの中にも、加工の効果の中にも、単純集計だけでは見えてこない偏りが存在し得るということには注意すべきである。

本来の標本抽出計画に近づけた加工データは、クロス集計を通した場合、センサスデータとの差は単純集計の比較以上に大きなものになりうるということがわかった。では、このようなデータにおいて、どのようにすれば国際比較というものが可能であると言い得るであろうか。

5.3 数量化3類を用いた国際比較

調査対象国の「調査文化」の影響を減らすために加工データを作成し、各質問項目の各回答カテゴリの単純集計や、属性項目の一部でのクロス集計の結果を、回収データやセンサスデータと比較してきた。ここまでの結論は、個々の質問項目では歪みが見えなくても、多変量レベルでは様々な予測しがたい差が生じているかもしれない、ということであった。

ただし、そのような差が存在するにしても、回収データと加工データが、他の国のデータとの比較においてある程度の安定した関係を示しているならば、その意味での国際比較は可能と考えることもできよう。そこで、北京、香港の回収データと加工データを、「東アジア価値観国際比較調査」の対象国、日本、上海、韓国、台湾、シンガポールと、同じ調査票を用いて同時期に行った中国本土の杭州市と昆明市の調査データ(鄭, 2005)と合わせ、それらの国・地域間の関係性を、林の数量化3類(林, 1992)を用いて比較してみたい。

まず、先述の性別と年齢階層、それに国・地域の項目を合わせて分析し、その結果を国・地域のみ表示したのが図6である。

北京、香港の回収データと各加工データの位置関係がかなり近いことが見て取れる。上海と北京が接近しており、上海と北京の各データを区別することは難しいが、他の国や地域との比較は十分可能ではないだろうか。

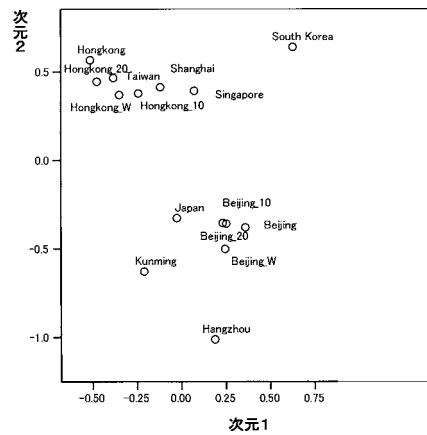


図 7. 全属性項目における国際比較. 性別, 年齢階層, 学歴, 職業, 世帯収入と国・地域を数量化 3 類で分析し, 国・地域のみをプロットした.

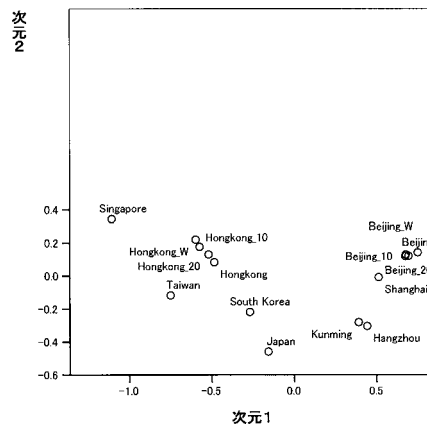


図 8. 社会的階層, 宗教感情における国際比較. 社会的階層と宗教感情(「神や仏」, 「死後の世界」, 「靈魂」, 「悪魔」, 「地獄」, 「天国や極楽」, 「宗教上の罪や罰」がある・存在すると思うか)と国・地域を数量化 3 類で分析し, 国・地域のみをプロットした.

さらに, 各国・各地域のデータで共通する全属性項目, 性別, 年齢階層, 学歴, 職業, 世帯収入(学歴, 職業, 世帯収入は各国の事情を加味して 3 段階に階層分けしている)を分析した結果が, 図 7 である.

この場合は, 香港, 台湾, 上海が接近しているが, これらを一つのクラスターと捉え, 他のクラスターと比較して特徴を捉えることはできる.

では, 通常の質問項目では, どうだろうか. 通常の質問項目の回答カテゴリーの内, 北京と香港の回収データと各加工データにおける差の絶対値が共通して大きいもの(4%以上)は, 問 6 の「社会階層」(5 段階に分けて自分の家族が所属すると思う階層を聞いている)を聞いた質問と, 問 11 の宗教感情(「神や仏」, 「死後の世界」, 「靈魂」, 「悪魔」, 「地獄」, 「天国や極楽」, 「宗教上の罪や罰」のそれぞれについて, ある・存在すると思うかどうかを聞いている)であった(取り上げた項目の詳細は, 吉野, 2004c, 2004d, 2006b を参照). これらと各国・各地域とを分

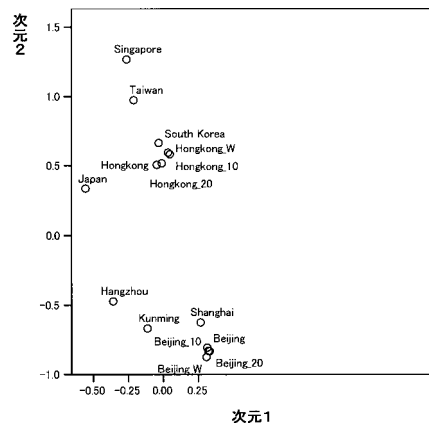


図9. 全質問項目における国際比較. 属性項目を除く全質問項目を数量化3類で分析し, 国・地域のみをプロットした.

析した結果が, 図8である.

この場合, 加工データと元の回収データの集まりと, 他の国・地域との関係性がかなり明確になっている. さらに全質問項目に対して行った結果が図9である.

国・地域を比較した全体の構図の中で, 北京と香港の回収データと各加工データの差異は, 国・地域間の差異と比較してかなり小さいこと(図中で位置が重なっている)が確認される. つまり, 個別の項目や, クロス集計した差異の比率にある程度の違いが見られたとしても, 多くの項目を同時に含めた多変量解析で国際比較を行う際は, 回収データと各加工データの差の影響はほとんどないということである.

6. 多様体としての文化

北京2002調査データ, 香港2002調査データは, 標本抽出方法に問題があり, それに対して3つの方法による加工を試み, 回収データと比較してきた. 単純集計における差異は標本抽出誤差の範囲に収まり, センサスデータと単純集計を比較した際でも大した差は見られなかったが, センサスデータをクロス集計し, それと回収・加工データを比較した際には, ある程度の差が存在した. しかし, 数量化3類を用いて多次元的に国際比較を行った場合, 回収データと各加工データとの差は分析に際してはほとんど影響がなく, かなり安定した結果が得られるということが分かった. (勿論, センサスデータと比べるだけでは, 母集団と回収データの意見分布の差という本質的な問題に答えたことにはならない(吉野, 2002).)

それでは何故, 標本抽出方法の差異があっても, 国際比較においてはデータの安定性がみられるのであろうか. 調査においては, 常に統計的無作為標本抽出を厳格に遂行することが求められるが, それが不要ということになりはしないであろうか.

もし, 調査の目的が, 国際比較ではなく, 単一の母集団(北京や香港といった一地域)における意見分布を導き出すことであるならば, 本調査のデータには偏りがあるかもしれない. 前節では, データに当初の標本計画に近づける加工を施して比較したが, 加工によって母集団(センサス)データから離れていく場合もあることを, クロス集計の結果から示した. 本論文では擬似的であったが, 理想的な状況は別にして, 標本抽出方法が違えばデータに違いが生まれうるということである. もちろん, 属性(性別や年齢)の偏りが意見分布に反映する場合, 例えば,

年金に関する質問を若年労働層と高齢の年金受給者層に対して聞く際に、回収データの年齢分布が母集団の年齢分布と比べて偏っている場合などでは、回収データからの推定値と真の値にはある一定の把握可能な偏りが出るということがありうる。このようなときには、第3節で批判した「母集団の属性に基づくウェイト補正」が正当化される。しかし、一般の調査では、どのような質問がどの層の人々に特徴的な反応をもたらすかは、あらかじめ明らかになっていない。それ故に、補正が正しく行われるか不明である「ウェイト補正」は行わず、回収データをそのままの形で取り扱うのである。

では、国際比較の場合はどうであろうか。我々は、「意識の国際比較」を可能にするために、「文化の多様体解析」という方法論に立つ。「文化の多様体解析」において、文化現象とは、各国・各地域が持つ文化的コアのようなものを反映しているのではなく、文化の調査者・分析者・科学者による見方、切り取り方の違いによって、異なる形で見えてくる多層的な構造を持つ「多様体」とみなされる。多層性の意味するところは、調査文化の違いに基づく調査方法ごとのデータの相違が見えるレベル、ある文化内の意見構造が捉えられるレベル、ある文化と他の文化との違いを判別できるレベル、各国・各地域をいくつかのクラスターに分けて特徴付けられるレベルというものが存在するということである。これらのレベルは、どれかが正しいというのではなく、あくまでも分析者の位置によって、分析を通して示される意見分布や各国・各地域の配置が変わってくるということである。本稿の結論は、各国・各地域を分類するレベルによっては、「調査文化」や調査方法の違いにもかかわらず、データの安定性が見受けられるということである。もちろん切り取り方によって、分析結果を恣意的に操作してはならない。そのためには「自らの調査文化」と「対象国の調査文化」の差異を意識し続け、データの分析時に「調査文化の差異」が強く出てくる場面と、それを無視できるほど大雑把な捉え方をしていない場面とを混同しないようにする必要がある。

7. 日本における調査文化の変容

これまで日本においては、比較的完全な住民基本台帳や選挙人名簿による標本抽出が可能であり、他国と比較すれば理想に近い世論調査ができる調査文化があった。しかし現在は、「統計的無作為抽出方法」が遂行できないほどの「調査文化の変容」を迎えている。回収率が戦後の90%台から80%台に落ち、問題化してきたのは1970年代(杉山, 1986)であり、やがて1990年代には「70%」基準なるものが暗黙に仮定されてきた。これは回収層と未回収層の賛否の意見分布が6:4と4:6とかなり逆転しているも、回収率が70%以上であれば、観測される回収層での賛否の割合と真の賛否の割合が逆転することはないという考え方で正当化される(吉野, 2006a)。しかし、世論調査を装った犯罪への危惧や、2005年に施行された個人情報保護法に対して名簿閲覧を実質上規制する地方公共団体も多くなるなか、国の機関ですらも、いつでも直ちに調査が遂行できるとは限らなくなっている。このような「調査文化の変容」を受けて、日本でも閲覧名簿に頼らない、いわゆるエリアサンプリングの採用が検討され始めている。(実際のところ、中国2002調査におけるエリア・サンプリング方法を公表後、各方面から詳細な説明や指導が求められた。)

50%以下の回収率のデータを「日本人全体の世論」と言えるのであろうか。第8回(1988年)日本人の国民性調査委員会・委員長、水野欽司が同調査の結果を受けて「調査の時代は終わった」と述べたとおり、本当に統計的標本抽出理論に基づく社会調査が机上のフィクションになってしまったのであろうか。この問題に対して、調査関係者が皆、緊急に解決策を模索しているところである。残念ながら、ここで完全な解決策を提示することはできないが、その解決に向けての試行錯誤への示唆として、4つの議論を提示することで、本論文のまとめに代えよう。

まず第一に、経験的には以下のようなノウハウが調査研究者には知られている。実験調査において、回収率60%で止めた回収データ(A)と、それにさらに調査期間を延長し、「不在」や「拒否」となった計画サンプルにさらに繰り返し接触と調査協力を試みて、回収率を70%まで向上させた回収データ(B)との比較において、 $(B-A)$ はAとかなり異なる分布であっても、B全体とAとは、多くの項目でほとんど同じ分布になることが知られている(吉野, 2002)。勿論、数学的に極端なケースを想定した反例を作ることはできるが、通常の世論調査で考えられるような状況では、上記のようになることが知られている。数学的に常に成立することが証明できるわけではないので、書籍や論文としては明示されることはなかったが、この種のノウハウは、回収率の上昇と情報量の逓減率(回収が進むにつれて、回答分布が収束すること)に関する理論の模索を示唆している。非回収層に関する研究において、土屋(2005)は、回答分布を推定することの困難さを指摘しているが、この考え方に同意している。

第二は、標本抽出理論に関する。従来の調査は、「調査毎」に「標本抽出理論」に基づく推定値を取り扱ってきたが、戦後長年にわたり同じ母集団(例、日本の成人全体)に対して、数々の調査を遂行し、結果を蓄積してきた。それらの多大の蓄積は、各々の調査結果のみならず、標本抽出誤差に関する情報をも提供する。場合によっては、単発の調査としてではなく、同じ調査票や調査項目の時系列的調査として、標本抽出誤差推定理論を考えることはできないであろうか。

数学的に一般的な証明はできないが、経験的には、「同調査機関」による一般世論調査の同じ調査票における同項目の回答分布は比較的安定しており、時系列的に変化がある場合でも、ほとんどの場合、緩やかな変化を示すことが多い。突発的な事件に敏感に一時的反応をする場合は除いて、経験的には、標本抽出誤差の幅と比較し時系列的変化はかなり安定している。これは、従来のように無作為に抽出するような理論を想定するだけでなく、人々の集団心理、世論の形成や伝播プロセスを考慮して、回答分布パターンの安定性を保証する理論を構築すべきことが示唆されているのではないだろうか。実際、まだ実用化されているわけではないが、Tourangeau et al. (2000)に見られるように、限定された条件の下では種々の心理プロセスを考慮した理論が試行されている。

第三は、「回収率」についてである。これは一般の社会調査については必ずしもいえないが、少なくとも政府やマスコミの「世論調査」について、従来の「回収率」の概念や定義の変更に関係する。従来は、調査協力を拒む人からは回答が得られないので、それは拒否率(非回収層)の中に入れられ、回答データの枠外に考えられていた。また戦後長きに渡り、2005年末くらいまでの政府の世論調査では、「政府(国)による」調査であることを伏せて民間調査会社が代行していた。このような方法が始まったのは、戦後、世論調査が開始された時期に、人々が戦前の思想調査と政府による弾圧を連想して、「本音」を言わないことがあったためらしい(日本民族学協会, 1952, p. 56)。しかし、2005年度中に、ある民間調査会社が日本銀行や内閣府の全国標本抽出調査において、代替サンプルの使用など、調査手続きの不正を行っていたこと、実際の回収率が50%台、あるいはそれ以下に低下していることが判明した。(但し、ここでも、その50%台のデータと、元の代替サンプルを含むデータにおける意見分布の差は項目毎には、ほとんど大差がなかったと事後報告されている。)

この事態を受けて、当局と筆者の一人である吉野らとの現今の社会状況を勘案する相談の中で、実験比較調査を試行しながら、「政府(国)の政策立案に国民の意見を反映するための基礎情報収集としての世論調査」であることを明示し、調査を遂行する方向が示唆された。その後、実験調査の結果、回収率が60数%になり、その方針で今後の世論調査が遂行されることとなったらしい。

明確に「政府の政策立案のため」と称した調査に、なおかつ拒否する回答者は、その態度自

体がその者の回答を表していると考えられないであろうか。世論調査は英語で public opinion poll (「投票」の意味あり)とも表す。実際の選挙でも投票に行かぬ人々も多い。しかし、投票率が低いからといって選挙が無効になることはなく、投票しない人は、他の人々の投票結果を受け入れる(委任)という態度を示しているのと同じになる。あるいは、せいぜい、当該の選挙争点に一定の態度を示すために「投票しない」という行動もありうる(90年代初期の無党派層)。いずれにせよ、無投票も投票行動の一つである。それと同様に、世論調査で回答拒否も「回答」の一部として組み込むことができないであろうか。

この1~2年の回答拒否の状況が政治的問題よりも、先にも述べたとおり、調査を装う犯罪への恐れや個人情報への過敏な反応からきている状況では、直ちには、上記の考えは首肯し難いかもしれない。しかし、調査主体と回収データの秘密保護についての十分な信頼性が担保できれば、「拒否」の持つ重要な情報(属性との関連など)も活用できるのではないであろうか。(Tourangeau et al., 2000の9章では、調査拒否、特定項目の回答拒否について、調査内容との関連を検討している。)

最後に、マルチ・モードとミックスド・モードとの調査法について触れておこう。松田(2006a)は、現在の調査回収率低下の問題に対して、調査モード(面接、電話、インターネット)毎に補足できる回答者層が異なることを利用して、ミックスド・モード調査の利用を提案している。しかし、実態調査と異なり、意見調査では同じ回答者ですら、異なるモードでは異なる回答をすることも多く、矛盾が生じることは知られている。どのモードの回答が正しいという問題ではなく、どのモードで調査するとどのような回答分布が得られるかという全プロセスが貴重な情報なのである。その意味で、それをマルチ・モード調査と呼び、異なる調査モードのデータを総合して1つの調査結果とするようなミックスド・モード調査とは弁別されるべきである。実際、松田(2006b)では面接調査と郵送調査において、生活満足度のようなありふれた質問項目でも、66%対34%と43%対56%のような回答分布の大きな逆転が生じていることを報告している。これでは、2つのモードのミックスの仕方でも如何様の結論でも出しかねない危惧がある。

また、内閣府政府広報室(2006)の実験調査のインターネット調査(調査研究者の用語として、厳密にはインターネット・モニター・アンケートと呼ぶべき)において、回答分布は(標本抽出法による)面接調査と比べて、生活満足度や一家団欒の満足度が著しく低い傾向を確認している。それでは「内閣支持率」が実態よりも低く出てしまい、政府の調査としては適正ではあるまい。いずれにせよ、安易にミックスド・モード調査に頼るのではなく、各調査手法やモードの特性を十分に把握し、それ自体を貴重な情報として取り扱うことが肝要である。

以上、完全な解決策には遠いが、調査研究者がそれを目指して進んでいくための議論のきっかけになるようなことがあれば幸いである。

謝 辞

本研究は、日本学術振興会からの科学研究費補助金・基盤研究A(課題番号14252013、代表・吉野諒三)「東アジア価値観国際比較調査—「信頼感」の統計科学的解析—」及び基盤研究A(課題番号18252001、代表・吉野諒三)「環太平洋価値観国際比較—東アジア周辺諸国の「信頼感」の統計科学的解析—」の一部として遂行されたものである。関係各位に深く感謝いたします。

参 考 文 献

- ESOMAR/WAPOR (2007). Guide to opinion polls, <http://www.esomar.org/uploads/pdf/ESOMAR-Codes&Guidelines-OpinionPolling-v5.pdf>.
- 林知己夫(1984). 『調査の科学』, 講談社ブルーバックス, 東京.
- 林知己夫(1992). 『数量化』, 朝倉書店, 東京.
- 林知己夫(2000). これからの国民性研究—人間研究の立場と地域研究・国際比較研究から計量的文明論の構築へ—, 統計数理, **48**(1), 33–66.
- 林知己夫(2001). 『データの科学』, 朝倉書店, 東京.
- 林知己夫, 吉野諒三, 鈴木達三, 林 文, 釜野さおり, 三宅一郎, 村上征勝, 佐々木正道(1998). 『国民性七か国比較』, 出光書店, 東京.
- 林 文, 山岡和枝(2001). 『調査の実際』, 朝倉書店, 東京.
- 本田則恵, 本川 明(2005). インターネット調査は社会調査に利用できるか, 労働政策研究報告書, **17**.
- 今井正俊(1996a). 朝日新聞世論調査半世紀の歩み(上), 朝日総研レポート, **122**, 94–106.
- 今井正俊(1996b). 朝日新聞世論調査半世紀の歩み(中), 朝日総研レポート, **123**, 108–120.
- 今井正俊(1997). 朝日新聞世論調査半世紀の歩み(下), 朝日総研レポート, **124**, 103–116.
- Kish, L. (1992). Weighting for unequal P_i , *The Journal of Official Statistics*, **8**, 183–200.
- 松田映二(2006a). いま世論調査が直面する壁, よろん, **97**, 37–40.
- 松田映二(2006b). 郵送・インターネット比較調査で世論調査の可能性を探る, 朝日総研レポート, **198**, 190–220.
- 内閣府政府広報室(2006). 訪問面接調査とインターネット調査の比較について, <http://www8.cao.go.jp/survey/sonota/h17-houhou/h17-houhou.pdf>.
- 日本民族学協会(1952). 特集「社会調査」, 民族学研究, **17**(1), 1–98.
- 杉山明子(1986). 社会調査の光と影, 言語生活, **411**, 28–34.
- 高橋正樹 編(2004). 科学史と科学者, 行動計量学, **31**, 107–124.
- Tourangeau, R., Rips, L. J. and Rasinski, K. (2000). *The Psychology of Survey Response*, Cambridge University Press, Cambridge.
- 土屋隆裕(2005). 調査不能者の特性に関する一考察, 統計数理, **53**(1), 35–56.
- 山岡和枝, 李 相侖(2004). 国際比較調査データの安定性についての検証—2003年度韓国と台湾における「健康と文化調査」および「東アジア価値観国際比較調査」データの比較—, 行動計量学, **31**(2), 125–135.
- 吉野諒三(2001). 『心を測る—個と集団の意識の科学—』, 朝倉書店, 東京.
- Yoshino, R. (2002). A time to trust—a study on peoples' sense of trust from a viewpoint of cross-national and longitudinal study on national character—, *Behaviormetrika*, **29**(2), 231–260.
- 吉野諒三(2002). 調査環境の中で, より信頼性を確保する標本抽出法とは何か, 平成13年度世論調査に関する調査研究・報告書, 66–72, 内閣府大臣官房政府広報室.
- 吉野諒三(2003). 日本における国民性研究, 『国民性論』(アレックス・インケルス 著, 吉野諒三 訳), 447–470, 出光書店, 東京.
- 吉野諒三(2004a). 科学的「世論」調査法の価値, 市場調査, **259**, 4–13.
- 吉野諒三(2004b). —特集—国民性を学際的に見る「国民性」の調査—計量的文明論の確立に向けて—, 学際, **12**, 6–12.
- 吉野諒三 編(2004c). 東アジア価値観国際比較調査—2002年日本調査報告—, 統計数理研究所研究レポート, No. 91.
- 吉野諒三 編(2004d). 東アジア価値観国際比較調査—2003年度中国調査(北京・上海・香港)報告—, 統計数理研究所.
- 吉野諒三(2005a). 東アジア価値観調査—文化多様体解析(CULMAN)に基づく計量文明論の構築へ向けて—, 行動計量学, **32**(2), 133–146.

- 吉野諒三 (2005b). 富国信頼の時代へ—東アジア価値観国際比較調査における信頼感の統計科学的解析—, 行動計量学, **32**(2), 147–160.
- 吉野諒三 (2006a). 今世論調査が直面する壁—「歴史」と「理論」と「実践」—, よろん, **97**, 31–36.
- 吉野諒三 編 (2006b). 東アジア価値観国際比較調査—総合報告書—, 文部科学省科学研究費補助金研究・基盤研究(A)研究成果報告書.
- Yoshino, R. (2006). A social value survey of China: on the change and stability in the Chinese globalization, *Behaviormetrika*, **33**(2), 111–130.
- Yoshino, R. and Hayashi, C. (2002). An overview of cultural link analysis of national character, *Behaviormetrika*, **29**(2), 125–141.
- 吉野諒三, 林知己夫, 鈴木達三 (1995). 国民性の国際比較調査の為の質問文の作成—翻訳のプロセスを中心として—, 行動計量学, **22**(1), 62–71.
- 鄭 躍軍 編 (2003a). 日本・中国の国民性比較のための基礎研究—中国北京市における意識調査, 統計数理研究所研究レポート, No. 89.
- 鄭 躍軍 編 (2003b). 日本・中国の国民性比較のための基礎研究—中国上海市における意識調査, 統計数理研究所研究レポート, No. 90.
- 鄭 躍軍 編 (2005). 日本・中国の国民性比較のための基礎研究—中国杭州市と昆明市における意識調査, 総合地球環境学研究所, **1**.

On the Stability of Public Opinion Data of Chinese Value Survey
with Respect to Sampling Methods
—A Note for the Development of Cultural Manifold Analysis
(CULMAN)—

Akira Horoiwa¹, Ryoza Yoshino^{2,3} and Yuejun Zheng⁴

¹National Institute of Science and Technology Policy,
Ministry of Education, Culture, Sports, Science and Technology

²The Institute of Statistical Mathematics

³The Graduate University for Advanced Studies

⁴Research Institute for Humanity and Nature

The main objective of this paper is to present an aspect of our research methodology for cross-national survey, Cultural Manifold Analysis, in the investigation on the stability of response data of China 2002 Survey carried out by the cross-national survey committee of the Institute of Statistical Mathematics. The main focus is on the reliability of response data of Beijing and Hong Kong. We assumed those data to be collected by three-stage random samplings: sampling of survey points proportionally to the ratios of population in the first stage, sampling of households at each of the selected points in the second stage, and sampling of a respondent at each of the selected households by a sort of birthday rule or the Kish method in the third stage. The outcomes, however, suggested that each of the final samples might have been made to be proportional to the ratios of population as if the selected sampling points were the total population, i.e., the selection probability of the numbers of households at each selected point might have been double.

We investigate the impact of probability sampling by comparing the originally collected data and the modified data that we made by re-sampling the same number of respondents at each of selected points (which we may assume to be closer to the correct probability sample). The result confirmed the stability of those data, i.e., there was no significant difference between the two sets of data in both univariate tabulation and multivariate analysis (Hayashi's quantification method III). Finally, some comments are provided for the future development of practical sampling theory in public opinion polls.