

「抽出による調査」と「割り当て調査（クォータ法）」

統計数理研究所
領域統計研究系

教授 鈴木 達 三

1. 序にかえて……日本の抽出調査（確率標本調査）の概要……

現在わが国で実施されている世論調査は、標本規模、調査対象者集団、対象地域、補助資料の利用状況等により、幾つかの形態はあるが、全国規模の調査における標本設計は基本的にはほぼ共通のものになっていると考えられる。

たとえば、広報室の世論調査報告書から調査の概要を抜き書きしてみると

- 1 調査の目的（調査ごとに異なる 略）
- 2 調査項目（略）
- 3 調査対象
 - (1) 母集団 全国20歳以上のもの
 - (2) 標本数（3,000、5,000、10,000人）
調査によって異なる
 - (3) 抽出方法 層化2段無作為抽出法
- 4 調査時期（調査期間は概ね10日間）
- 5 調査方法 調査員による面接聴取法
- 6 調査実施委託機関（略）
- 7 回 集 結 果
 - (1) 有効回収数（率） ○○○○人
（○○.○%）
 - (2) 調査不能数（率） ○○○人
（○○.○%）

—不能内訳—
転居○○、長期不在○○、一時不在○○、
住所不明○○、拒否○○○、その他（病
気など）○○

8 性・年齢別回収結果

	性・年齢	標本数	回収数	回収率
男	20～29歳			
	30～39歳			
	⋮			
	60～69歳			
	70歳以上			
	計			

	性・年齢	標本数	回収数	回収率
女	20～29歳			
	30～39歳			
	⋮			
	60～69歳			
	70歳以上			
	計			

等となっている。

ここで、本項に関連のあるのは3、(4、7)である。

まず、3の(3)抽出方法について多少専門的になるが、報告書の標本抽出方法から概要をのべておく（この例は本年1月実施の標本数3,000人のもの）。

1 地域層別

全国の市区町村を、都道府県を単位とする11地区に分類し、地区内では都市規模（大都市では各都市ごと、人口10万以上の市、人口10万未満の市、町村の14分類）により分類し第1次層を計44層とした。

2 第1次抽出単位の層化と抽出
(層化)

第1次抽出単位は昭和60年国勢調査の調査区とし、第1次層内を昭和60年国勢調査時の調査区特性により層化した。

(調査地点数)

第1次層の推定母集団数(昭和63年3月31日現在の20歳以上人口)の大きさにより標本3,000を各層に比例配分し、調査地点の標本数が11~16となるように調査地点数を決めた。

(抽出)

第1次層からの調査地点(調査区)の抽出は無作為に抽出した。

3 調査地点における対象者の抽出は、調査地点の範囲内(町、丁目、番地等を指定)で標本となる対象者が抽出できるように抽出間隔を算出し、住民基本台帳により等間隔抽出法によって抽出した。

ここにあげた調査事例では全国の176市町村、210調査地点が抽出されている。

ここまで調査の概要をのべると、諸外国における調査との比較を考える時の諸点がはっきりしてくる。と同時に、われわれの日常実施している調査の基盤が多くの資料によって支えられていることがわかる。

本項の意図は、上にのべたような現行の調査を念頭において「確率抽出による調査(full probability sampling「確率標本法」)」について

1 日本と諸外国(特にアメリカ)との違い

2 現行調査の問題点

を考え、さらにこれらに関連して考えられている「割り当て調査(quota samplingクォータ法)」について

- 1 アメリカにおけるクォータ法の実際
- 2 日本におけるクォータ法の理解のされ方などをのべることにある。

2. アメリカにおける確率標本企画の一例

現在、アメリカで全国規模の確率標本企画を立て、実際に調査活動をしているところの1つは、ミシガン大学の社会調査研究所(ISR, Institute for Social Research)の一部門の調査研究センター(SRC, Survey Research Center)である。ここがシカゴ大学の全米世論調査センター(NORC, National Opinion Research Center)と共同で企画・実施した全国標本計画について、概要を示す

Heeringa, Steven G., Judith H. Connor, and Doris C. Darrah, 1980 SRC National Sample, Design and Development, February 1986, Institute for Social Research, Ann Arbor, Michigan.

1980年の国勢調査の結果資料を利用した全国標本計画が最新のものである、アメリカでは10年ごとに人口および住宅に関する国勢調査が実施され、その最新の結果が公表されるごとに、国勢調査局、NORC、SRCや他の主要機関が全国標本計画を更新してきたが、全国規模の確率標本を計画し維持・管理するのに多大の費用がかかるので今回はSRCとNORCが共同で計画し維持・管理することになった。実際の作業は1982-83年におこなわれた。

国勢調査の資料を利用して標本企画を立てるところは日本と共通しているが具体化されてくると相違が目立ってくる。

全国を東北部、中北部、南部、西部の4地域と大都会かそうでないかで分類していくのは日本と共通であるが、この区分は国勢調査資料の区分に

よっている。

<第1段の層別>

層別は州と郡(county)をもとにしているが、まずニューヨーク(近郊を含む以下同じ)、ロサンゼルス、シカゴ等の16の大都市圏はそれぞれ独立の層となり(16層)、つぎにSMSA*地区(Standard Metropolitan Statistical Area)……日本の人口集中地区の区分のような、都市的性格の強い郡、又は郡の集合体……を地域別に45層に分け、それ以外(Non SMSA's)の地区では地域別に23層あわせて、第1次層として84層に分割した。

* 国勢調査局による区分、説明は、例えば「統計学持論」、日本放送出版協会 昭和61年のP97~P101参照

大都市圏、SMSA地区、それ以外の地区に分けた人口および居住家屋数(Occ. HU's Occupied Housing unit's)は、表1ようになる。

表 人口と居住家屋数(大分類)

Classification	Total	
	人口 Pop (千人)	居住家屋数 Occ. HU's (千戸)
Total 全体		
No. Primary Areas 1980 Counts	226,505	84
大都市圏 self-representing		
No. of Areas 1980 Counts	61,269	16
それ以外 Non self-representing		
No. of Areas 1980 Counts	165,236	68
SMSA地区 SMSA's		
No. of Areas 1980 Counts	108,136	45
それ以外の地区 Non-SMSA's		
No. of Areas 1980 Counts	57,100	23

<第1段の抽出>

全国の84層のうち、16の大都市圏はそれぞれが第1次抽出単位として抽出され、残りの68層からは、郡(あるいは郡の集合体)を第1次抽出単位とし、各層から(郡の)大きさ(居住家屋数)に比例した抽出確率で1単位を抽出した。

<第2段の抽出>

第2段抽出における抽出単位は「エリア・セグメント(area segments)」といわれ、大都市圏およびSMSA地区では国勢調査の調査区(census blocks)、あるいはセンサス・ブロックの集合体を、また、それ以外の地区では国勢統計区(Enumeration districts, ED's)からなり、それぞれに含まれる居住家屋数の大きさ(が50以上になるように調整して)に比例した抽出確率で抽出される。

第2段階の抽出における抽出単位数は大都市圏ではその大きさ(第1次抽出単位の大きさ)により、(ニューヨーク49、ロサンゼルス38、シカゴ35等と異なるが…)計314、それ以外の層では各層から12、計(12×68=)816あわせて1,130単位を抽出した。抽出は1980年国勢調査のサマリー・テープから電算機により行った。

ここまでで、調査地域の抽出が終わり、つぎに地域に出かけて、地域の地図(見取図)、家の配置図を作成し世帯のリストアップをすることになる。今回はSRCとNORCが1,130を折半して実施した。

以上のようにして全国標本が出来上ることになる。日本のように調査対象とくい違いの殆んどない抽出台帳(住民台帳、選挙人名簿)が存在しないため、調査地点における抽出作業の基礎として、地図および住宅配置図が欠かせないことになり、それぞれの抽出段階においても人口と居住家屋数

とが必ず出てくる。

世帯のリスティングは、計画全体の調査数を見込みその需要には応じられるように考えているが、人口の社会移動のはげしいところでは、転入（新築）率および空室率が推定以上になったりするので、リストの維持・管理は容易ではない。

したがって、全体としてみたとき、調査全体の中に占める標本抽出関係の費用の割合がかなり大きく（聞くところによれば調査規模にもよるが標本数1,500として調査費用の1/4見当といわれる）、また、調査地区を調査ごとに抽出し直すことなど不可能であるから、頭初の標本設計時における層別方法等々が日本より重要になるものと考えられる。

しかし、逆に考えれば、一度標本計画を立て地

域を抽出してしまうとそのシステムを今後10年間固定的に利用する形になるので、調査実施を考えると日本とは別の考え方が重要になる。

これらの点は日本で通常考えている標本調査の企画、調査の実施における留意点とは異なるので慎重な研究が必要になるといえる。

また、日本と異なる点は、抽出の各段階で家屋を考えているので、世帯単位の調査ではなく個人対象の調査の場合には調査対象に指定された家屋からの個人抽出をすることになり、最終の結果推定にはこのウェイトを考えることになる点も注意が必要である。

実際の確率標本による調査計画の一例として1986年の大統領選挙の際の調査の標本企画と調査実施状況の結果を示す。

表 2 Original Sample Design Specifications and Assumptions and Actual Sample Design Outcomes. 1986 NES Post-Election Survey

	Original Specifications and Assumptions	Actual Outcome
Completed interviews	2400	2176
Response Rate	.72	.677
Eligible sample households	3333	3215
Occupancy/Eligibility Rate ^{*1}	.87	.835
Final sample HU listings	3831	3848
Sample growth from update ^{*2}	1.03	1.035
Sample listings from frame	3719	3719

*1 Eligibility (.97) x occupancy (.90).

*2 Since the updating process produces about a 3% increase in sample lines over the count selected from our National Sample system, the update inflation factor was set at 1.03.

3. 日本における現行調査の問題点

最近の問題点は大きくって2つある。1つは現在では、一応解決されたとはいっても問題の残る抽出名簿のことである。周知のように、住民基

本台帳の閲覧がダイレクトメール屋などの関連やプライバシー問題などのため、事実上大きく制限されてきた。これに伴って、名簿を見せていただくという立場上、名簿の質について問題を提起す

ることが出来にくい状況になってきている。このため①最近時の名簿が閲覧出来るかどうか不明確、②名簿の仕様が抽出地点の地域の細部とマッチしているかどうか不明確等の点が問題になってくる。この他、名簿閲覧の制限が強いため、調査実施上からみて、その情報があればフィールドワークが容易になると考えられるような各種の補助情報が得にくくなっている。

これらの諸点は、つぎにのべる第2の問題点である「調査票回収率の低下」の問題とも密接に関連しているので慎重な検討が必要である。

第2の問題は「回収率の低下傾向」であり、この対策を考えることであろう。これについては、昨年度行われた内閣広報室の「今後の世論調査の在り方についての研究」において若干のべられているので、ここでは指摘するだけにしておくが、このことが、つぎの「割り当て法」を考える一つの動機になっているものと思われる。（これはF. Kingが「クォータ・サンプリング」の冒頭にのべていることを思い起こさせる）

King, B.F. (1983) "Quota Sampling" in *Incomplete Data in Sample Surveys*, Vol. 2: Theory and Bibliographies, eds. W. G. Madow, I. Olkin, and D. B. Rubin. New York: Academic Press.

4. アメリカにおけるクォータ法の実際

① 現在実施されているクォータ法の概要

現在実施されているクォータ法 (quota sampling) は、比較的短い調査期間で計画通りの調査完了数が得られ、くり返し訪問にかかる出費もない方法であり、Sudman (1967) が「潜在的なバイアスの最も少ない形式 "probability sampling with quotas" と名付けている方法である。

この方法は、地域の層別、地域の抽出から調査地点のブロックの抽出までは確率標本法（2参照）と全く同じ手順を用いている。調査地点が抽出された後、家屋のリストを作成し、世帯や個人を無作為抽出するかわりに、調査員は特定の属性項目や個人的判断に従い指示された割り当てをみたとすようにする。

割り当ての指示は、通常は管理運営上ごく限定されているが、調査抽出地域を直接とりまく最小地域における最近時の国勢調査のデータによってきめられている。

たとえば、調査員はある市のブロックの特定の街角の世帯から、指示された道順でつぎつぎに訪問し、男女成人5人づつを訪問、面接調査するまでつづけるように指示される。不在や拒否は無視して次の家に行く……

以下、実際の調査機関で実施されている例をあげる。

Sudman S (1967). *Reducing the Cost of Surveys*. Chicago, Aldine

② NORCのクォータ法

◎調査地点は、1978年のNORCの全国確率標本のために抽出されたブロック（あるいは擬似的にブロック化された地点）と同じ（これは通常の層別多段確率比例抽出法と同じ 2参照）

◎各調査地点当り5人の回答者が得られるように考える。出発点はブロックの中で選定され、以後調査員は曲がりくねったブロックに沿って行くように指示される。

National Opinion Research Center (1978). *General Social Surveys, 1972-1978: Cumulative Code-book*. Chicago: National Opinion Research Center

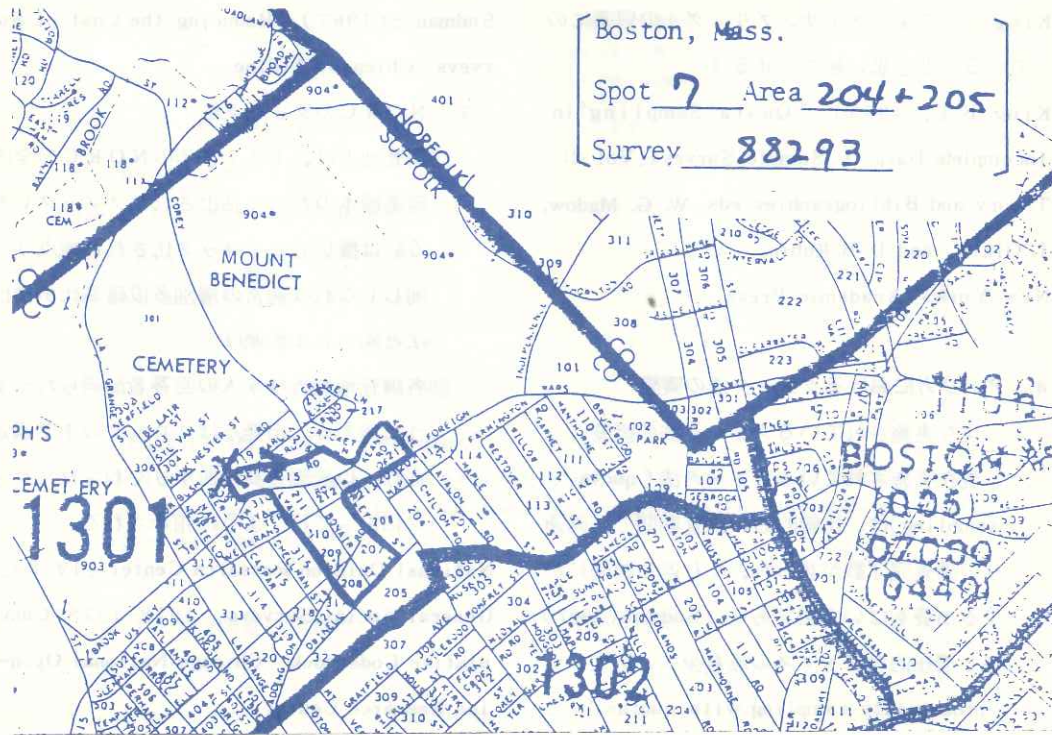
◎調査員は平日は午後3時以降、土曜・日曜

は一日調査するようになってい

◎性別の割り当てに加えて男性では年齢の区分-30歳が分割点を女性では従業上の地位-勤めているかどうかで割り当てる。(この考えは、若い男性と雇用されている女性は調査の際容易にはつかまらない傾向があるからである)

◎回答者集団への割り当ては、ブロックの出発点を含む国勢調査区あるいは小地域に対する性、年齢、従業上の地位(雇用かどうか)の構成比による(これらの属性をクロスした統計資料が公表されている最小単位の国勢調査の地域統計データによる)

③ ギャラップ調査機関のクォータ法
調査地点の抽出:クォータ法をはじめ地域の抽出はNORCの場合と同様な確率的抽出
個人の抽出:それ以降の回答者抽出の手順



はギャラップ調査機関のオムニバス調査の概要にある通り*。

すなわち、各抽出された地域ブロック、あるいはセグメントにおいてランダムに抽出された出発点は、その調査地点の調査員地図の上に示される(図参照)。この地点から出発して指定された道順に沿って訪問する家をつぎつぎに選定する。このようにつぎつぎに世帯を抽出し、その地点に割り当てられた数の訪問面接調査を完了するまで調査をつづける。居住している住宅あるいは世帯において、個人の抽出は

- i) 在宅している18歳以上の男性のうち最も若い人に
 - ii) もし男性がいなければ、18歳以上の女性のうち最も年齢の高い人に
- 面接する。

世帯内における調査対象者のこの選定方法

は性別にみた年齢構成が母集団の年齢構成と比較して近くなるように経験的に開発されたものである。

世帯内で在宅している人を選ぶこの方法はランダムではないが系統的であり客観的である。

訪問面接:これは一般に成人が在宅している時を考慮し、週末および平日では女性は午後4時以降、男性は午後6時以降に実施される。不在の場合あるいは該当する成人対象者がいないときはつぎに行き、くり返し訪問はしない。(Politz-Simmonsの在宅率に対する修正をして偏りを減少させる方式をとっている)

* The Gallup Organization, Inc (1972),
The Design of the Sample for Trends,
The Personal Interview Omnibus Survey,
Princeton, N. J.
また、説明は前出'統計学持論' P97~P101参照。

参考までに、ギャラップ調査の調査員地図の一例を示す。(前頁参照)

④ クォータ法についての問題点*
クォータ法に関する根本的なものは「回答者の選出機構が不明確であり、通常のランダム抽出のように、目的の母集団に対して適切な推定をするのは難しい」ということである。

この問題は、確率標本法における、(母集団からの標本)抽出枠(フレーム)の欠陥の問題や、無回答、調査不能による偏りの問題と同様なものではあるが、確率サンプルの場合は、抽出枠を慎重に検討することによって、また、フィールドワークを十分にやり、調査

結果の追跡検討を十分やることによって、結果に含まれる偏りがある一定の限界内に収めることは可能であろう。ところがクォータ法の場合には、どの程度全体をカバーしているかという程度も不明確であり、その結果生じる偏りの程度もはっきりしないというのである...

King, B.F. (1983) 前出「クォータ・サンプリング」参照
クォータ法の偏りの主な原因のいくつかは、たとえば

a) 回答者の在宅率の違いによる偏り
くり返し訪問しないのでクォータ法では在宅率の低い居住者に対する偏りが考えられる。(これまでに示した手続きにより、この欠陥を修正するように計画されているがうまくいっているかどうか不明)。しかし、これは割り当てに利用する属性項目を増せば減少する。

b) 割り当てに利用する変数が不十分なこと
より多くのコントロール変数を利用して、調査員に割り当てさせれば、目的とする母集団に対してより一層偏りの少ないサンプルにすることができる。しかし、こまかい割り当てをすればますますこの割り当て通りの属性をもつ回答者を得ることは困難となり、確率標本方式より相対的に有利であるとみられてきた調査スピード、調査費用の利点は失われる可能性が高くなる。(複雑すぎる割り当ては調査員を混乱させるだけでメリットはなくなる)

c) 調査員が回答者を選択するときの偏り
ギャラップの方式の年齢に関する指示以外では、調査員は指示された割り当て層の中から誰を選ぶか全く自由である。したがって、どの人に面接するかをきめるとき各種

の偏りが考えられる。(この選択は、回答者の態度や性格に深く関連している可能性がある)

d) 層の大きさに関する情報が不正確であること

各抽出地点における、クォータ法の属性割り当ては、その地域をカバーする直近の最小単位における属性の分布に従ってなされる。——通常これは国勢調査の調査区統計資料等になる——

この資料が抽出調査地域と完全にはマッチしていないとか、国勢調査時から日時が経過しすぎているとかの程度により、偏りが生じる可能性がある。

また、割り当て法では、調査地点当りのサンプル数を通常同一に固定しているので、調査地点が代表する層(地点抽出のもとになった地域層)の人口分布の大きな変動を反映できないようになっている。

これらの影響および他の偏りの可能性のあるもの等は評価するのが困難である。

しかし、これらのことが、確率サンプルの方法をその通りに実施する調査機関の能力不足によって生じる欠陥より悪いということは明らかではない。(特に面接調査に対する抵抗が強まるという状況では、何度も再訪問をくり返すとか協力をうながすとかしても回収率はそれほど高くないから同様の問題に直面する)

⑤ クォータ法と確率抽出法の実際の調査結果に基づき経験的な比較について(アメリカの事例)

よく知られているようにトルーマンとデューイが争った1948年のアメリカ大統領選挙の選挙予想で、割り当て法による各調査機

関の予想は失敗した。このときの社会科学研究委員会の報告^{*1}は、主要な調査機関で用いられてきた回答者の選出方法——割り当て法——に大きな関心を払い、予想の失敗にはサンプル選択方式だけではなく、誰に投票するかきめてない層や選挙期間の終盤での投票者の態度変化の予測の方式も重大な関連があるとした(割り当て法では低い学歴層、地方に離れて住む人が少ない傾向があることが判明している)。報告書には確率標本法と割り当て法の両結果を併記して前者がよい結果を示しているところもあるが、結論として、割り当て法を止めるようにするよりも、サンプル抽出の方式および調査員の訓練についてもっと注意を払うべきであると報告している。

この予想の失敗により、調査機関はより一層慎重な方式手順をとるようになったと考えられている。

その後、StephanとMcCarthy^{*2}(1958)は「割り当て法」と「確率標本法」による調査結果について比較をして報告している。両方よく合っている場合もあればそうでない場合もあり、必ずしも「割り当て法」はよくないということではなかった。結論的にみて、「割り当て法」による重大な偏りは屢々割り当てに利用する社会経済的な属性の制御に関連している。したがって、正確さに欠ける調査員判断による社会経済的階層の決定法にかえて、調査地域を含む小地域の地域情報を利用するのがよいとした。またSudman(1967)^{*3}はNORCの実験調査の結果、両者よく似ていたが「割り当て法」のサンプルは1人あるいは2人家族の世帯が少なくなっていた。これから、将来「割り当て法」に世帯人数の条件をつけた実験調査を計画すべきであると提

案している。——これは現在まで実施されていないところである)。

*1 Mosteller, F., Hyman, H., McCarthy, P., Marks, E. S. and Truman, D. B.

(1949), The Pre-Election Polls of 1948, Bulletin 60, New York: Social Science Research Council

*2 Stephan, F., and McCarthy, P. J. (1958), Sampling Opinion. New York: Wiley

*3 前出

最後にNORCが毎年実施している一般社会調査(G. S. S.)においておこなった「割り当て法」と「確率標本法」の比較について述べる。

これは1975年と76年のG. S. S.において抽出された地点を折半し、その一方は再訪問付き確率標本方式、他方はblock内を一軒一軒訪問していく「クォータ」方式で実施された。

この両方式は同じ地域に同時期に住む人に対して実施されたので、比較のとき生じる可能性のある多くの変数がコントロールされている。Stephenson^{*1}(1979)は多数の属性項目および態度項目に関する比較の結果を報告している。

一般的に認識されているような小人数の世帯に対する偏り(これは他の比較には大きな影響はないとみられる)はあるが、これ以外に二つの方法における重大な差異はない。しかし、「クォータ法」では雇用されている男性の比率が少ないこと、これは、「この偏りを補正しようとして導入している年齢のコントロールが有効に働かなかったことを示していると考えられる」(King)。

*1 Stephenson, C. B. (1979), Probability sampling with quota: An experiment. The Public Opinion Quarterly 43(4): 477-496.

⑥ 「クォータ法」の偏りを推定する方法

クォータ法の偏りを推定し、それを修正しようという試みがStinchcombe^{*1}(1971)、King(1974, 1985)にある。

Stinchcombeは再調査を実施し、その過程で、調査の回答者になりやすいかどうか、個人の見つかる分布を推定し、これを利用して偏りを推定する方式を考えた。

Kingは「クォータ法」と「確率標本法」の場合について、偏った測定のパイゾモデルを適用して一つの比較方式を提案した。^{*2}

さらにKing(1985)は「クォータ法」が理論的に正当化される条件について検討を試みている。^{*3}

以上、Kingの「クォータ・サンプリング」に沿ってアメリカにおけるクォータ法の実際についてみてきた。

*1 Stinchcombe, A. (1971) The Computation of Bias in Quota Samples. (Unpublished)

*2 King, B. F. (1974) "Surveys combining probability and quota methods of sampling" in Proceedings of the Social Statistics Section, American Statistical Association, 351-356.

*3 King, B. F. (1985) "Surveys Combining Probability and Quota Methods of Sampling" Journal of the American Statistical Association Vol180, No392. Theory and Methods 890-896

5. 結びにかえて

以上のべたことを土台にして、これらによって、日本におけるこれまでのクォータ法の理解のされ方、今後の運用の仕方等を考えてみるのも重要であろうと思う。

最近では、県単位の選挙予想の際における電話調査の回答者選択に「クォータ法」が利用される

ケースがみられる。

まだ、日本においては実験的あるいは試行錯誤の段階であるから、どうすべきというよりも、調査資料をつみ上げて慎重に検討することから始めなければならないと考える。

