

## RDD サンプリングの理論と実践

島田喜郎  
株式会社日経リサーチ

キーワード:

電話調査、RDD、稼動局番フレーム、電話帳準  
拠フレーム

要約:

本論文は RDD サンプリングに関連する実務上の諸問題について、理論的な根拠に基づいた対処方法を考察する。主な論点は以下の通りである。

- RDD では稼動局番フレームや電話帳準拠フレームという電話番号リストから、単純無作為抽出や系統抽出で電話番号の標本を作る(2 節)。
- 抽出した標本には非使用の電話番号が多数含まれるがスクリーニングによって実査開始前にその番号を処理済にすることができる(3 節)。
- 世帯用番号の密度は地域によりばらつきがあるが、抽出フレーム全体に均一の抽出率を適用すると自動的に密度差は調整される(4.1 節)
- 電話調査では調査地点の抽出をせず、一段階の抽出で直接世帯を選ぶことができる(4.2 節)。
- 抽出フレーム内を局番順にソートして系統抽出することにより暗黙の層化ができる(4.3 節)。
- 個人調査では世帯内の適格者から一人を無作為抽出し、集計時に抽出率を調整する(5 節)。

また、最後に今後の研究課題について論じた。

## 1. はじめに

本論文の目的は、ランダム・ディジット・ダイアリング(RDD)を使って実際に調査の標本設計を行う実務家が遭遇する諸問題について、理論的な根拠に基づいた対処方法を論じることである。RDD サンプルングは、その理論的な枠組みが抽象的で、想像力を働かさないと意外な落とし穴に落ちることがある。本論文のねらいは、標本設計者の想像力を助け、重要なポイントでの意志決定に論理的な根拠を与えることである。

取り上げた問題は、電話番号リストに相当する抽出フレームの作り方、抽出フレームから標本を抽出する方法、使われていない電話番号や業務用電話番号のスクリーニング、世帯用番号密度の地域差の扱い方、多段抽出や層化の考え方、RDD を使った個人調査、などである。

## 2. RDDの抽出フレームと抽出方法

### 2.1. RDDの原理

RDD サンプルングの母集団は固定電話の加入世帯(以後「電話世帯」という)である。世帯を対象とした調査の本来の母集団は、電話加入の有無に関わらず世帯全体であるが、RDD サンプルングでは、わずかながら存在する非加入世帯に目をつぶり、電話世帯をもって母集団の代用とするのである。

NTT の統計によると、「加入電話」と「ISDN」を合わせた固定電話の「住宅用契約」の件数は2002年に4,265万であった。これを国勢調査の世帯数4,706万(2000年10月)で割ると加入

率は90%を少し上回る。この他に「事務用契約」で住宅用として併用されているものがあり得るので、それを考慮すると、固定電話の加入率は90数%に達しているものと思われる。また、2000年の固定電話の普及率が96%であるという調査結果もある(佐藤, 2002, p.200)。このように高い普及率を持つ電話システムを、標本抽出と対象者へのアプローチの両方の手段として活用しようとするのが RDD 電話調査である。

世帯の母集団リストは世帯主の住所氏名で構成されるのに対し、電話世帯の母集団リストは世帯用電話番号で構成される。電話帳掲載率が百パーセントであれば、電話帳をもって電話世帯の母集団リストとすることができるが、最近では掲載率が低下して網羅性が失われている。RDD サンプルングの目的は、電話帳に掲載されていない電話世帯が多数存在するという状況の中で、掲載番号にも非掲載番号にも等しい抽出確率を与えて世帯用番号の標本を作ることである。その目的のために、世帯用番号を完全に(あるいはほぼ完全に)包含し、かつリスト可能な抽出フレームを定義する。この抽出フレーム内の電話番号から無作為抽出を行って電話番号の標本をつくる。得られた標本がRDD標本であり、このRDD標本の中で世帯用番号と判明したものを電話世帯の標本とする。

RDDの原理の中で最も重要なポイントは、抽出フレームには世帯用番号の他に、業務用番号も、使われていない番号も含まれるが、それらすべてに等しい抽出確率を与えることによって、世帯用番号同士が互いに等しい抽出確率を与えられることである。その結果として、RDD 標本の電話番号のうち世帯用と判明した

ものを電話世帯の標本とすることが可能になる。

RDD 標本のように、いったん得られた標本の中に不適格な固体が含まれ、各個体にコンタクトして初めて適格な固体を確定できるという状況は決して特殊なものではない。たとえば成人のうち、過去一年間に海外旅行をしたものの標本を得たい場合、まず成人の標本を作り、それから選ばれた個人にコンタクトして、旅行の経験を確認して適格な個人を確定する。

## 2.2. RDDの抽出フレーム

### ○稼動局番フレーム

上記の抽出フレームを単純に具体化したものが、表1の上半分に示した「稼動局番フレーム」である。稼動局番フレームは、固定電話システムで稼動中の局番、すなわち現に存在している市外局番と市内局番の組み合わせ（たとえば012-345-xxxxの数值部分）をすべて集めたものである。この抽出フレームでは局番のリストが作られるが、実際に抽出の対象となるのは局番内の個々の電話番号である。局番内に末尾が0000から9999の1万個の電話番号が列挙されていると考えれば、局番のリストは同時に電話番号のリストであるともいえる。

この抽出フレームには現在の固定電話システム内に存在し得るすべての電話番号が含まれる。当然、世帯用電話番号もその中に完全に包含される。稼動局番フレームに含まれる局番の数は全国で約1万9千、この中の電話番号の数は1億9千万に上る。全国の世帯用番号の件数は約4千3百万であるので、稼動局番フレーム中の世帯用番号の比率は20数%である。このフレームから無作為抽出した電話番号

が世帯に当たるヒット率も平均20数%ということになる。

### ○電話帳準拠フレーム

稼動局番フレームの世帯ヒット率の低さを改善するために考えられたのが、表1の下半分に示した「電話帳準拠(list-assisted)フレーム」である。電話帳準拠フレームは、電話帳(list)の情報を使って作成する。電話帳に掲載されている世帯用番号が1以上、あるいは一定件数以上含まれるバンクを有効バンクとして、有効バンクだけを集めて抽出フレームを作る。

バンクは局番の中を一定の大きさのブロックに分けたもので、012-345-67xxのように局番の先頭から百位までの数字が同じ100個の電話番号のブロックを百位バンクと呼ぶ。千位バンクや十位バンクも考えられるし、局番は万位のバンクである。

局番の中の電話番号は、世帯用、事業所用、非使用のものが入り混じっているが、それぞれが特定のバンクに比較的集中して存在する。その結果、電話帳掲載の世帯用番号が1以上または数件以上ある「有効バンク」には世帯用番号がある程度以上存在すると考えられるし、掲載番号が1つも無い「無効バンク」には世帯用番号がまったくないかわずかしかないと考えられる。有効バンクだけを集めて抽出フレームを作ることにより、抽出フレーム内の世帯用番号の密度が大幅に高まり、標本の世帯ヒット率を高めることができる。

電話帳準拠フレームでは有効バンクのリストが作られるが、実際に抽出の対象となるのはバンク内の個々の電話番号である。百位バンクであればバンク内に末尾が00から99の百個の電話番号が抽出対象として存在する。バ

ンク内にその電話番号が列挙されていると考えれば、バンクのリストは同時に電話番号のリストであるともいえる。

有効バンクを集めて抽出フレームを作ると世帯ヒット率は高まるが、その程度はバンクの大きさや、有効と判定する掲載件数の基準によって違ってくる。土屋・前田(2003)は千位バンクを使ったが、松田(2002)は百位バンクを使っている。米国では、最近ではほとんどの場合百位バンクが使われている。百位バンクの方が世帯ヒット率の改善度が大きいからであろう。また、RDD 標本抽出の専門会社が百位バンクのデータベースを使っているという事情もある。百位のバンクを有効と判定する掲載件数の基準値をみると、米国では、厳密さを要求する政府機関の調査では1、効率を重視するマーケティング調査では2ないし3が使われることが多い。

電話帳準拠フレームは、世帯用番号がすべて非掲載であるバンクを排除してしまうので、別名脚きり(truncated)フレームとも言われる。脚きりのため電話世帯のカバレッジが完全でないことが問題とされることがある。日本では脚きりについての認識は比較的甘く、「Bank もれの問題」は止むを得ない(松田 2002、p.83)という程度に考えられているようだ。その結果、脚きりされたバンクの世帯用番号が「全体としてどのくらいの率になるか実際に確かめた例を見たことがない」(鈴木 2003、p78)というのが実情である。脚きりによるバイアスについては、Brick, Waksberg, Kulp and Starer (1995)や Giesbrecht, Kulp, Starer (1996)の研究がある。彼らは、掲載番号1以上の百位バンクを使った場合、1)脚きりされる世帯は電話世帯の3-4%にすぎない、2)脚きりされた世帯と抽出フレー

ム内の世帯に有意な差がない、という二つのポイントを確認し、脚きりによるバイアスは重大でないと断定した。米国で重要な調査に電話帳準拠フレームが使われるようになったのは、これらの研究以降である(島田 2003)。電話帳掲載状況は、時代によって変化し、国によって違いがある。ほぼ十年前の状況をベースにした研究が現在の米国に当てはまるか疑問があるし、そのまま日本に当てはめることができないことは確かだが、問題の取り組み方としては参考にすべきである。

### 2.3. RDDの抽出方法

抽出フレームから電話番号を抽出するとき、すべての電話番号を物理的に列挙する必要はない。局番やバンクを列挙して、その中に0000-9999 や 00-99 の末尾をもつ電話番号が列挙されているかのように想定すればよい。想定上の電話番号リストから RDD 標本を抽出する方法は、一般の標本抽出と変わるところが無い。

#### ○単純無作為抽出

稼動局番フレームから電話番号を単純無作為抽出するには以下のようにする。A市に稼動局番が30局あるとすると、このフレーム内に30万件の電話番号が存在する。準備として局番に0から29までの連番をふる。0から299,999の範囲の乱数を発生させる。乱数で236,125番目の電話番号が選ばれると、その電話番号は連番23の局番に属する。その局番が047-473なら、局番内の6125番目の電話番号は047-473-6125である。

電話帳準拠フレームから単純無作為抽出す

る方法も上記とほとんど同じである。A 市に電話帳掲載の世帯用番号を 1 件以上もつ百位バンクが 1500 あるとする。このフレーム内の電話番号は 15 万件。バンクに 0 から 1499 までの連番をふる。0 から 149,999 の範囲の乱数を発生させる。乱数で 81,435 番目の電話番号が選ばれると、その電話番号は連番 814 のバンクに属する。そのバンクの電話番号の百位以上の桁が 047-472-25 なら、バンク内の 35 番目の電話番号は 047-472-2535 である。

A 市の電話世帯数が 10 万とすると、稼動局番フレームから抽出した RDD 標本の平均世帯ヒット率は 10 万/30 万で 33% である。電話帳準拠フレームの RDD 標本の平均世帯ヒット率は 10 万/15 万で 67% である。

A 市の例で示した抽出方法はそのまま全国に拡大できる。日本全国の稼動局番の数は約 1 万 9 千であるので、稼動局番フレームで使う乱数の範囲は 0 から 1 億 9 千万になる。電話帳準拠フレームでは、1 以上の掲載番号をもつ百位バンクの数が 70 万とか 80 万に達するだろう。その中の電話番号の件数は 7~8 千万件になる。

#### ○系統抽出

系統抽出を行う場合、ごく普通に系統抽出をすると、末尾の数字が 00、01、10 のように特別の用途に使われそうな番号が繰り返し選ばれる心配がある。それを避けるために少し工夫が必要だ。抽出フレームをインターバルの大きさのブロックの集合とみなし、各ブロックからランダムに一つの電話番号を抽出する。これは「ゾーン分け抽出法」といわれるものに相当する。

### 3. 電話番号のスクリーニング

電話番号のスクリーニングとは抽出後の RDD 標本に対して機械的な処理を行って、使われていない番号を判定したり、業務用の可能性が高い番号を判定したりすることである。これにより RDD サンプリングの最大の問題点である実査の非効率性を大幅に改善することができる。これは、標本抽出後、実査を開始する前に行う重要なプロセスである。

#### 3.1. スクリーニングのタイプ

電話交換機は、ダイヤルされた電話番号を受け取ってから、実際に交換作業を開始する前に、その番号の性状を示す電子信号を発信側に返してくる。その電子信号によって使われていない番号を判定できる。コンピュータを使って RDD 標本の番号をダイヤルしてこの処理を行うと、多数の電話番号を「非使用番号」としてふるい分けできる。

一方、RDD 標本に選ばれた電話番号を事業所電話帳とつき合わせて、「事業所用番号」である可能性が高いとしてマークすることができる。事業所用番号は代表番号だけが電話帳に掲載されているのが普通であるから、すべての事業所用番号をマークできるわけではない。また、飲食店や床屋など、事業所用と掲載されていても世帯用に併用されているケースが考えられるので、マークされたものすべてを直ちに事業所用と確定しないほうが安全だろう。

#### 3.2. スクリーニングした番号の処理

スクリーニングの結果、非使用と判定された番号や確実に事業所用であると判定された番

号でも、それを RDD 標本から排除することはしない。それぞれの電話番号の処理分類 (disposition) を「非使用」や「事業所用」と確定し、その番号に対する実査業務が完了したものとみなす。

非使用や事業所用と判定した番号を RDD 標本から排除し、それに代わる番号を同じ局番やバンクから再抽出する処理をすると、そうした不適格番号の周辺の抽出率が高まってしまう。たとえばあるバンクで世帯用番号が 1 件だけ存在し、他はすべて非使用番号であったとする。このバンク内のいずれかの番号が抽出されると、再抽出が繰り返されバンク内の唯一の世帯用番号が必ず選ばれることになる。このような処理は避けなければならない。

事業所用と確定できないが、その可能性が高いとマークされた番号には、平日の昼間にダイヤルすることにより効率よく処理を進めることができる。

### 3.3. 抽出フレームとスクリーニングの関係

機械的なスクリーニングで、非使用と判定される番号の比率は、使用した抽出フレームにより異なる。全国の稼働局番フレームを使った日経リサーチの経験では、非使用番号の比率は 64% に達する。前田・土屋(2001) は一都三県で千位バンクの電話帳準拠フレームを使って、自動判定された非使用番号が 26% であると報告している。

稼働局番フレームでスクリーニングの効果が大きいのは、このフレームがもともと非使用番号を多数含んでいるのだから当然である。しかし、稼働局番フレームにとって、スクリーニングは単に効果が大きいという以上の意味があ

る。もともと稼働局番フレームは、極端に世帯ヒット率が低いいため現実の調査に使えらる人は少なかった。しかし、コンピュータを使ったスクリーニングによって 1 件数円の費用で使われていない電話番号を判別でき、残った電話番号の世帯ヒット率が実施可能レベルになることが分かると、この抽出フレームのもつ長所が際立ってくる。それは電話帳準拠フレームに比べて 1) 脚きりが発生することがなく、電話世帯を完全にカバーできる、そして 2) 電話帳データベースを準備する必要がないことである。稼働局番フレームでも、局番のデータベースは必要だが、局番データベースは 2 万件未満の規模だが、電話帳データベースは 3 千万件以上の巨大な規模になる。

稼働局番フレームから直接無作為抽出した RDD 標本を実用的に使っている例は少ない。非効率という先入観があるためである。しかし、機械的スクリーニングによって 64% の番号を実査開始前に「非使用番号」と確定できれば、世帯ヒット率の分母が 0.36 倍に縮小し、理論的なヒット率は 2.7 倍の 60% 前後にまで達する。これは決して非効率とはいえない。日経リサーチでは 2002 年の 9 月から、日経新聞社の定期世論調査をこの方法で実施している。

## 4. RDDにおける地域の問題

### 4.1. 世帯用番号密度の地域間格差

抽出フレーム内の世帯用番号の比率(以下では世帯番号密度という)は地域別にばらつきがある。RDD サンプリングに取り組む標本設計者は、このばらつきをどう扱うべきか悩むことがある。たとえば、都会と田舎の世帯用番号密

度の違いをどのように扱ったらよいか(松田 2002、p.83)、また、市町村別の世帯用番号密度をどのように推定するのか(佐藤 2002、p.194)という問題である。

しかし、実は RDD の基本原理を単純に適用すれば世帯用番号密度を推定する必要は無いのである。すなわち、世帯用番号密度が高い部分でも低い部分でも、「世帯用番号であるなしに関わらず、抽出フレーム内のすべての電話番号に等しい抽出確率を与える。」その結果として、世帯用番号が互いに等しい抽出確率で選ばれることが保証される。

上記の原理を実感として納得するためには、具体的な状況を考えて見るのが効果的である。表2によって、世帯用番号密度の異なる二つの市について考えてみよう。X市とY市の電話加入世帯数がそれぞれ10万世帯で、市内局番の数がX市は20局、Y市は40局である。局番内の世帯用番号密度はX市がY市の2倍になる。この密度差に関わらず、両市に共通の抽出率を適用する。1局番につき100の電話番号を無作為に抽出するとX市では2,000、Y市では4,000の電話番号をもつRDD標本が得られる。この標本の中の世帯用番号件数の期待値はX市、Y市ともに1,000件で、両市の電話世帯数に比例した適切な結果が得られる。地域により世帯用番号密度が大きく異なるにもかかわらず、一定の抽出率を適用したことによって適切な標本が得られたのである。

地域による世帯用番号密度のばらつきを無視できるということは、地域別の世帯用番号密度を知る必要がないことを意味している。もともと世帯用番号密度の情報は入手不可能であり、入手不可能な情報が不必要であることによって、RDDサンプリングが現実的な方法といえる

のである。

#### 4.2 RDDと地点抽出

訪問面接調査などで使われる多段抽出は、世帯や個人の対象者を選ぶ前に地域の選択を行うことにより、対象者を限られた数の調査地点にまとめて、必要な費用や人員を現実的な制約の範囲内に抑えることを目的に行われる。しかし、電話調査では対象世帯が限られた数の調査地点に固まっている必要はない。電話は番号をダイヤルするだけで、日本の隅々まで瞬時にアクセスすることができる。そのため、一段階の抽出で直接全国の任意の電話番号もつ世帯に到達できるのである。抽出設計に地域を選択する段階を含めることは、誤差の限界を大きくするという損失こそあれ、特別の利得があるとは思えない。しかし、日本のRDDサンプリングの先駆的な実施例では、多段抽出を行っていると推察される例がいくつも見られる(田中 1995 p.73、谷口 1996 p.54、松田 2002 p.88)。それぞれ設計上の理由があるとは思いますが、「材料の性質に素直に沿った調理法」を取るもののほうが有利ではないだろうか。(有名な Waksberg (1978)の方法は、二段抽出を使っているが、一段目の抽出で世帯用番号を含まないバンクを排除するという明確な目的がある。Waksberg法は日本での利用例が非常に少ない。城川(2003)がその稀有の例と言える。Waksberg法は手順が複雑で時間と労力を要するため、米国でも最近あまり使われなくなっているという事情があり、この論文では取り上げなかった。Waksberg法の説明、また米国における最近のRDD法の動向については拙著、島田(2003, 2004)を参照されたい。)

#### 4.3. 地域の層化

前節で述べたとおり、面接調査では抽出の一段階として調査地点の抽出を行う。そのため、市区町村を都市規模や産業構造で層化することが標本抽出に不可欠のステップとなる。「限られた数」の「調査地点(固体の集合)」が、無作為抽出の結果、大都市部に偏ったとか、農村部に偏った場合その影響が大きいからだ。RDD では「多数」の「個々の固体」を一段階の抽出で選択できるため、面接調査の場合ほど地域の偏りを心配する必要は無い。

しかし層化が調査精度の向上に結びつくことは確かなので、効率的に層化を行う方法があればそれを取り入れるべきだ。土屋・前田(2003, p.106)は、「電話番号の層化を行うために電話帳に頼らざるを得ない」としているが、電話帳データベースを使うよりはるかに低コストで層化の効果をあげる方法もある。適切に配列したリストを使って系統抽出を行うと、層化と同等の効果を挙げることができることは良く知られている。RDD でも、局番をまず市外局番と市内局番でソートしてから系統抽出を行うと、Lepkowski (1988, p.82)が述べている通り、「暗黙の内に層化抽出を行ったことになる。」これは、簡単でかつ現実的な方法である。電話の局番は、厳密には地理的な境界と対応していないが、大まかに見れば地理的な境界によく対応している。そのため局番を番号順に並べて系統抽出を行うときめ細かな地域の層化を行ったのと同等になる。

#### 4.4. 区域外の番号を含む局番

電話局のサービスエリアが調査対象区域の境界を越えて広がっていることがある。選挙予

測調査などでは、区域を厳密に区切ることが必要になり、区域外の番号の処理に十分な注意が必要である。区域外の電話番号は「区域内」という適格条件に合致しない不適格番号で、使われていない番号や事業所用番号と同じと考えればよい。まず、区域外の番号を持つ局番(あるいはバンク)も、区域内だけの番号をもつ局番(バンク)と同じように抽出フレームに加える。RDD 標本の内、そのような局番(バンク)に属する番号にマークしておき、ダイヤルした相手が出た段階で、対象区域に属するかどうかを確認する。その結果、区域外であれば不適格として処理する。ある局番(バンク)が、独立してサンプリングする二つの区域にまたがっている場合、その局番(バンク)は両方の区域の抽出フレームに加える必要がある。

### 5. RDDによる個人調査

#### 5.1. 個人の抽出方法

RDD サンプリングを使って、個人を対象とした調査も実施できる。むしろ、個人調査のほうが断然多いだろう。個人調査では、電話が世帯につながったことを確認した上で、世帯内に対象者の条件に適合する個人がいるかどうかを確かめる。適格な個人が二人以上いる場合は、無作為に一人を選んで、その人に調査への協力を依頼する。無作為に個人を選ぶには、CATI に乱数を発生させたり、手元に適格者数別の乱数表を用意しておいて、適格者の年齢順に乱数番目の人を選ぶ。選ばれた人が電話に出ている人でないなら、当然その人を電話口には呼んでもらう。その人が不在なら、再度電話することになる。

乱数を使って厳密に対象者を選ぶとともに、ランダムに選ばれた個人を最後まで追いかけることが非常に重要だ。回答者の性別、年齢、職業などの構成が適切な比率に収まるのは、「乱数」があらゆるタイプの人を公平に選んでくれるからである。たまたま電話口に出た人を対象者にして調査を進めてしまうと回答者の構成はどんどんゆがんでいく。結果として属性別に回答者の数を管理しなければならなくなる。それは最早、無作為抽出ではなく、「クォーターサンプリング」になってしまう。

## 5.2. 抽出確率の調整

RDD サンプリングでは世帯が等確率で選ばれる。世帯から一人の適格者を選ぶ場合、個人の抽出確率は世帯内の適格者の人数に逆比例する。このとき発生する抽出確率の差は、集計時に世帯の適格者の人数で重み付けすることによって調整する。適格者の人数は、調査票の中の1つの質問として聞くが、適格者でなく「家族の人数」を答えてしまうことが無いよう十分な注意が必要である。

ついでながら、世帯が固定電話番号を二つ以上持っている場合にも抽出確率の差が発生する。番号を二つ持つ世帯は普通の世帯に較べて2倍の抽出確率を持つ。この場合は番号件数の逆数で重み付けする調整が必要である。電話番号の件数についての質問もまた注意を要する質問である。固定電話の件数を聞いているのに、携帯電話や親子電話の数まで含めて答えてしまわないだろうか。質問の仕方を工夫するとともに、上限値を設けるなどの対応も必要だろう。Massey and Botman (1988)は番号件数の上限値を2とした。

## 6. おわりに

### 6.1. まとめ

この論文では、さまざまな問題を取り上げた。それに一つの結論をつけるのは難しいが、あえて言えば、当たり前のことなのだが、特殊に見える RDD も還元してみれば普通のサンプリングだということだ。電話番号のリストを作り、単純無作為抽出や系統抽出によって電話番号の標本を作る。抽出フレーム内のすべての電話番号に等しい抽出確率を与えるのだから、抽出の時はすべての電話番号を平等に扱い、世帯用番号密度のばらつきは詮索しない。RDD サンプリングが普通のサンプリングと同じに見えてきたとき、RDD 調査をより適格に処理する基盤ができたといえるのではないだろうか。

### 6.2. 残された問題

実務上の重要な問題だがこの論文で扱うことを避けた問題がある。標本の不足・補充、重みつき集計などに関連する問題だ。これらの問題は RDD 調査に限ったものではないが、RDD では不確定や拒否が発生するステップが多いため、これらの問題の切実度は他のタイプの調査以上に高い。日経リサーチの世論調査では、標本の補充はしない、重み付け集計はしない(ただし、世帯の適格者数と電話件数調整はする)のが方針であるため、筆者はこうした問題についての経験が少なく、論じることを断念せざるを得なかった。

たとえば、重み付け集計についてみれば、RDD 調査といえば重み付け集計をするのが普

通のようだ(松田 2002 p.86)。重み付け集計をする理由は、全世帯とか全有権者のような本来の母集団を代表するデータでなければ、データの利用者には価値が無いと考えるからだ。これに対して、日経リサーチのように重み付け集計をしない立場は、回答を得られなかった人を代表するようなデータを作ることはできないというものだ。

重み付けでは、電話非加入者は、同じ属性をもつ加入者と同じ答えをしたはずとして処理され、調査拒否者は、同じ属性をもつ協力者と同じ答えをしたはずだと処理される。結局、調査に協力した人の回答がベースになっているので、重み付け集計をした結果としない結果が大きく異なることはまれである。もしその差が大きくなるようであれば、もともとの調査の標本構成がゆがみすぎているのである。

世論調査の世界では RDD 調査であるかどうかに関わらず「候補者の当落は予想できても、投票率の予想は絶対当たらない」というジレンマがある。調査で投票に行くと答える人の比率は、いつも実際の投票率よりも大幅に高いのだ。これは重み付け集計で調整できるレベルではない。原因は、回答者が建前で答えているとか、調査拒否者の多くが投票に行かないタイプの人であると考えられる。ともかく、母集団の真実の姿は重み付けをする、しないとはもつと別のところにあるのではないかというのが筆者の感想である。しかし、この問題について説得力のある議論をする準備はない。

筆者が論じることを避けた問題は、実務家にとっては避けて通ることのできない問題であり、今後の重要な、研究課題であると考ええる。

### 6.3. 謝辞

最後に、筆者の拙い文章を辛抱強く読んでいただいた編集委員および二人のレビューアーの先生に心からお礼申し上げます。誤謬や疑問点のご指摘に赤面し、適切なコメントに励まされながら、この論文を完成できたことに心から感謝します。

### 引用文献

Brick, J., Waksberg, J., Kulp, D., and Starer, A. (1995) Bias in List-Assisted Telephone Samples. *Public Opinion Quarterly*, Vol. 59, pp. 218-235.

Giesbrecht, L., Kulp, D., and Starer, A. (1996) Estimating Coverage Bias in RDD Samples with Current Population Survey Data. *Proceedings of the Survey Research Method Section, American Statistical Association*, pp. 503-508.

城川美佳 (2003) 医療・保険領域の電話調査 Waksberg 法. *行動計量学*, 58 号, pp. 111-119.

Lepkowski, M. (1988) Telephone Sampling Method in The United States. Groves, R., Biemer, P., Lyberg, L., Massey, J., Nicholls, W., and Waksberg, J., (eds.) *Telephone Survey Methodology*, John Wiley & Sons, pp. 73-98.

前田忠彦・土屋隆裕 (2001) 「日本人の国民性 2000 年度吟味調査報告 -電話・郵送・面接調査の比較-」統計数理研究所

Massey, J., and Botman, S. (1988) Weighting

Adjustments for Random Digit Dialed Surveys. 40-46.

Groves, R., Biemer, P., Lyberg, L., Massey, J., Nicholls, W., and Waksberg, J., (eds.) *Telephone Survey Methodology*, John Wiley & Sons, pp. 143-160.

松田英二 (2002) 朝日新聞社の RDD 調査について. 行動計量学, 56 号, pp. 81-89.

佐藤武嗣 (2002) RDD 電話調査. 林知己夫編「社会調査ハンドブック」朝倉書店, pp. 191-200

島田喜郎 (2003) 米国における RDD サンプルングの最近の動向. 日本行動計量学会第 35 回大会発表論文抄録集, pp. 126-129

島田喜郎 (2004) RDD サンプルング手法の比較研究. よろん, 93 号, pp. 45-55

鈴木達三 (2003) 電話調査についての一考察. 行動計量学, 58 号, pp. 73-91.

田中愛治 (1995) RDD 法による電話世論調査. よろん, 75 号, pp. 70-78.

谷口哲一郎 (1996) RDD 法の試行および問題点の検討. よろん, 78 号, pp. 51-64.

土屋隆裕・前田忠彦 (2003) 二種類の電話調査法の比較実験調査. 行動計量学, 58 号, pp. 93-107.

Waksberg, J. (1978) Sampling Methods for Random Digit Dialing. *Journal of the American Statistical Association*, Vol. 73, No. 367, pp.

表1 RDDの抽出フレーム

稼動局番フレーム	
定義	稼動しているすべての局番の末尾が0000-9999の電話番号
カバー率	電話世帯を完全にカバー
ヒット率	非常に低い
電話帳準拠(list-assisted)フレーム	
定義(百位バンクの場合)	掲載番号を一定数以上持つ百位バンクの末尾が00-99の電話番号
カバー率	脚きりが発生する
ヒット率	大幅に改善する

表2 世帯用番号密度と抽出率の関係

	X市	Y市
固定電話加入世帯	10万	10万
局番数	20	40
抽出率	100/局番	100/局番
抽出数	2000	4000
世帯用番号件数 期待値	1000	1000