

# 日本におけるマイクロデータ提供の現状

井出 満（大阪産業大学・経済学部）

## 1. マイクロデータの活用の必要性

現代の社会・経済状況は、複雑かつ不透明なものである。そのため、指定統計調査を実施する府省が集計・公表している統計表だけでは、複雑かつ不透明な社会・経済状況を明らかにすることが困難になっている。したがって、マイクロデータすなわち個票データのベースまで遡って分析する必要があるわけである。

もちろん、これらの分析を指定統計調査実施府省の統計関係部門あるいは行政・研究部門でも行うことができるかもしれない。しかし、知識や経験が豊富な学者、研究者などがこれらの分析を行い、これらの結果に基づいて行政に各種の政策提案することも、重要な役割分担であると思う。

諸外国では、マイクロデータの活用が積極的である。もちろん調査対象のプライバシーに関しては、誤差の付加、データの交換（スワッピング）などを行っている。しかし、わが国では、このような方法は、心理的な抵抗が大きい。

そこで、後述するマイクロデータの使用を閉鎖された場所、機関等で行うとか、あるいはリサンプリング、地域情報の広域化、トップコーディングなどの方法を採用し、マイクロデータの使用が可能になることが望まれる。

## 2. マイクロデータの活用の現状

### （1）1995年以前

後述する「統計情報活用のフロンティアの拡大」の研究が開始する以前のマイクロデータの提供は、指定統計調査を実施する府省が研究等の目的のために、学者、研究者などに依頼するケースのみであったといえる。すなわち、統計法第15条第2項の「統計上の目的以外の使用」<sup>1)</sup>により、総務大臣の承認を得て使用目的を公示してマイクロデータを使用することができるが、使用目的が公益性の高いこと、使用者が原則として公務員（税務職員および警察職員を除く）などの基準が厳しいため、前述のようなケースしかなかったわけである。

## (2) 「統計情報活用のフロンティアの拡大」の活動

文部省科学研究費補助金特定領域研究「統計情報活用のフロンティアの拡大ーマイクロデータによる社会構造解析ー」(代表:松田芳郎一橋大学経済研究所教授)が1996年~1998年にかけて実施された。特定領域研究参加者は、全国で約130人に達した。

この研究では、まず、研究参加者が研究のための集計表、関数などを事務局(一橋大学経済研究所)に提出する。目的外申請は、事務局で一括して行う。目的外の使用が許可されたものについて、参加者が集計のためのプログラムを作成し、そのプログラムを使用して事務局が集計し、その結果を申請者に渡し、参加者は、その結果を分析するという仕組みであるが、マイクロデータの活用の成果が大いに上がった。

しかし、参加者は、一切マイクロデータに手を触れなかったわけである。そのため、試行錯誤して説明変数を決めるような関数を求める場合、研究者自身がデータそのものを扱いたいとの要望が多く出たわけである。

## (3) 「マイクロ統計データ活用研究会」の活動

前述の特定領域の参加者をはじめ多くの研究者等は、目的外利用に対して政府の対応が可能になるまで、何らかの形で公開利用を継続してほしいという要望が強かった。この要請に応えるため、日本学術振興会の研究費補助金(研究成果公開促進費)の交付を受けて、マイクロ統計データ活用研究会(代表:当初は松田芳郎東京国際大学経済学部教授、現在は井出満大阪産業大学経済学部教授)が2000年~2004年の5年計画で発足した。なお、実際の活動は、1999年から始まっていた。

このプロジェクトでは、提供するマイクロデータは、使用許可が得られた統計調査の個票データから10分の1あるいは5分の1の部分標本データで、かつ秘密保護の観点から、当然ながら、世帯番号等の個体の識別子を消去するとともに、一部の調査項目を削除したり、調査対象の特定に結びつく可能性のあるデータは、例えば、世帯人員が7人以上の場合、「7人」に変換したりするなどの加工を行っている(このデータをリサンプリング・データといっている)。

また、今後のマイクロデータを提供する際、最も有力視されているリサンプリング・データについて、そのユーザビリティを検証することになっている。この検証は、公益性の高い研究ということができ、総務省統計局においても、この研究を高く評価し、指定統計調査の統計目的以外の使用を承認しているわけである。

なお、この研究会では、研究参加者は公募で行われ、目的外の申請およびリサンプリング・データの作成は、(財)統計情報研究開発センターが行っている。このリサンプリングの方法であるが、抽出率により調査市町村が特定できないように、また、研究参加者が容易に集計できるように、抽出率の逆数をウェイトとする確率比例系統抽出し、単純集計が可能になっている。

なお、目的以外の使用が許可された研究参加者は、リサンプリング・データを直接手に触れて集計することになっている。この集計後、全標本データを用いて、(財)統計情報研究開発センターで

同じ集計表あるいは関数などを集計し、リサンプリング・データのユーザビリティを検証している。

#### (4) その他の活動

総務省統計研修所では、総務省統計局が目的外利用を包括承認で求め、研究官室の客員教授あるいは客員研究官などの閉鎖された研修所の中で、マイクロデータの利用を行っている。

また、一橋大学では、経済研究所社会科学統計情報研究センターにマイクロデータ分析セクションが設けられ、マイクロデータのパイロットシステムの構築を行っている。

### 3. 今後の方向

#### (1) 統計審議会の「統計行政の新中・長期構想」の提言

情報通信技術の進歩により多様な集計が可能となってきたことに伴い、また、収集された指定統計調査の調査票が国民の共有の財産であり、多様な統計ニーズに対応して弾力的に統計の作成ができるよう、目的外の使用の運用をより弾力的にしてもらいたいとの要望が強くなってきた。

1995年3月10日に統計審議会が答申した「統計行政の新中・長期構想」では、「内外の研究者や国際機関等から、標本データ（簡票データから必要に応じて抽出を行い、地域区分や世帯番号等の個体の識別子を消去するなど個体の識別を不可能にしたもの）の提供を求める要望が高まっている。…………… このため、標本データの提供については、個体の秘密保護の担保方策を中心に、外国の制度及び提供例、国内外におけるニーズの実態、現行制度との関係、具体的な提供方策等について、おおむね2～3年を目途に専門的・技術的な研究を行う必要がある。」と述べられている。

この答申に基づいて、総務省統計局を中心に多くの検討あるいは研究が行われたが、残念ながら、マイクロ統計データの提供については、一定の方向を見出すことができずに推移している。

#### (2) 「統計行政の新たな展開方向」の申合せ

2003年6月27日に取りまとめた「統計行政の新たな展開方向」（各府省統計主管部局長等会議の申合せ）によると、「統計データの利用促進のための基盤整備」の章を設け、その一つとして、「匿名標本データの作成・利用」を掲げている。その基本方向として、「…………… 、匿名標本データの作成・利用について、秘密の保護を前提に、…………… 、実用化に向けて検討する。」としており、また、具体的方策として、「…………… 、必要に応じ試験的な匿名標本データの作成を行う。」としている。

いずれにしても、どのような形にしる、マイクロデータを学者、研究者などに提供されることが望まれる。また、学者、研究者などの卵である大学院生などに対しては、小さな標本（スワッピングしたデータでもよい）であっても構わないが、目的外使用の申請を必要としないようなモデルデータ

(レプリカデータ)の提供が切に望まれるところである。

**(参考条文)**

- 1) 第15条 何人も、指定統計を作成するために集められた調査票を、統計上の目的以外に使用してはならない。
- ② 前項の規定は、総務大臣の承認を得て使用の目的を公示したものについては、これを適用しない。

**(参考文献)**

総務庁統計局統計基準部監修『統計行政の新中・長期構想—統計審議会答申』 (財)全国統計協会連合会 1995年  
坂本信三『我が国の統計制度—統計の総合調整と統計法規の解説—』 (財)全国統計協会連合会 1999年  
松田芳郎編『統計情報活用のフロンティアの拡大—マイクロデータによる社会構造解析—』 (財)統計情報研究開発センター