

まえがき

従来から統計は集団現象を記述する資料とされてきた。このために、調査の結果得られるデータは基本的に集計表の形で提供され、また統計学も一方では集団現象の諸側面を統計データとして可能な限り正確に反映する統計調査論、また他方では与えられた集計結果表から集団現象に貫く統計的規則性（統計的法則）を析出するための方法の学問として展開されてきた。

集団現象を構成する個々の要素の中には規則性を反映する要素と統計的誤差に相当するものとが混在している。このうち統計的誤差、なかでも偶然誤差といわれる部分については、多数の事例を合計することで相互に相殺され、結果的に規則性のみがのこる。このような大数の原理が集団記述の学問としての統計学の背景になっている。

伝統的に統計が専ら集計量として提供されてきたのには、それなりの理由がある。なぜなら、最近でこそ高性能の計算機が大量情報の加工、処理を可能にするようになったものの、その普及以前は、調査の結果得られた大量情報の処理は、大規模集計機械を有する政府機関のみがなす業務であった。このような状況の下では統計利用者にとっては、統計の作成者である政府機関が提供する集計結果表は文字通り所与のものであり、利用者は、それからの読みとりあるいは部分的な組み替え集計によってそこに反映された規則性を発見するという形で統計を活用してきた。

統計のこのような利用形態は、大容量情報の加工処理を可能にする近代的な高性能情報処理技術の普及の結果大きく変貌をとげる。すなわち、統計利用者自身が政府統計調査等によって得られた個別データ（非集計データ）を独自に加工処理し、分析等に活用をはかることが少なくとも技術的には可能になった。統計作成者である政府機関が統計作成に当たって本来目的としていた統計の利用とは異なる目的への統計の利用は、一般に政府統計の二次的利用といわれている。既存の集計結果表の利活用といった従来型の統計利用に比べ、個別データの潜在的利用可能性は格段に大きく、それだけに二次的利用の中心が非集計個別データとなることは、現在の情報処理技術水準を前提すれば当然のことといえよう。

欧米のいくつかの国ではすでに 1960 年代半ばからマイクロデータと一般に呼ばれている匿名化された個別データが提供されてきた。それは、研究等の分野で広範に利用され、それまでの集計データだけでは明らかにできなかった事柄について、多くの有効な分析成果が出されている。本書の論文でも紹介されているように、欧米の他の多くの国でも、1980 年代から 90 年代にかけて、あらたにマイクロデータの提供に踏み切る動きが見られる。さらに近年では、途上国の中にも政府統計から匿名標本データを作成し、それを公共目的のマイクロデータ（Public Use Files:PUF）として公開する動きがあり、現在では、世界の多くの国で政府統計マイクロデータの提供が、統計作成機関の統計活動の中に構造的に組み込まれた観さえある。

わが国では、政府統計に関する個別データの提供は、指定統計の目的外使用制度の運用として行われてきた。しかしこのような提供方式に関しては多くの問題点が指摘されており、わが国の統計行政の将来の在り方を示した統計審議会の答申においても、その改善の

必要性が指摘されている。また 1990 年代半ば以降、わが国でも統計研究者を中心に個別データを用いた分析が行われるようになり、政府マイクロデータの作成、提供を求める声が強くなってきた。

このような動きを受けて、総務省統計局では、同局が所管する全国消費実態調査、就業構造基本調査、住宅土地統計調査等の大規模標本調査から標本再抽出により匿名標本データ（マイクロデータ）を試験的に作成し、現行制度の枠組みの中でそれを分析目的に提供している。本報告書の第 2 部は、これらのリサンプリングデータに基づく分析結果を収録したものである。なお、本書では、第 1 部として、わが国における統計学の展開の中にマイクロデータをどう位置づけるかといった統計理論上の問題、わが国におけるマイクロデータ提供の現状、さらにはマイクロデータ分析に関する基本的手法の解説をとりあつかった諸論文を収録することにした。

本書が、政府統計マイクロデータに対する広範な関心をよび、わが国における政府統計マイクロデータの作成、提供システムの実現に向けて貢献できれば幸いである。

2004 年 4 月
法政大学 日本統計研究所

